Majestic International Journal of AI Innovations
Volume 1, 2026, Article ID: CM2601105004
DOI: 10.65080/inprocess

CONTENT MAJESTIC PUBLISHER

OPEN ACCESS

Research Article

Hybrid CNN–LSTM Intrusion Detection Framework for Industrial IoT Security

Mushtaq Ali[1, *] Imad Ullah[1]

[1]*Riphah Institute of Informatics, Malakand Campus, Riphah International University Islamabad, Pakistan*

Article History

Received: 16 January, 2026 Revised: 27 February, 2026 Accepted: 12 March, 2026 Published: 24 March, 2026

Abstract:

Aims: The rapid adoption of the Industrial Internet of Things (IIoT) has increased the exposure of safety-critical industrial systems to sophisticated cyberattacks, necessitating intrusion detection mechanisms that are both accurate and operationally reliable. Traditional intrusion detection systems were shown to struggle with the complex statistical correlations and temporal dynamics inherent in IIoT traffic, often resulting in elevated false alarm rates and unreliable detection of evolving attacks. This study proposed a lightweight hybrid CNN–LSTM intrusion detection framework that jointly modelled correlated statistical traffic descriptors and their temporal evolution.

Methods: Convolutional layers were used to learn compact representations from high-dimensional IIoT traffic features, while a Long Short-Term Memory (LSTM) layer captured temporal dependencies associated with multi-stage and slow-rate attacks. The model was evaluated on a large-scale Industrial IoT intrusion detection dataset using imbalance-aware metrics, threshold analysis, and probability calibration.

Results: Experimental results demonstrated strong generalisation performance, achieving an F1-score of 0.968 and a ROC-AUC of 0.780 under controlled experimental conditions. Importantly, the proposed approach maintained a low and controllable false alarm rate across a wide range of decision thresholds and produced well-calibrated prediction confidences.

Conclusion: These properties indicated that the framework was suitable for real-time deployment in resource-constrained IIoT environments, where operational reliability and risk-aware decision making are critical.

Keywords: Industrial internet of things; intrusion detection system; deep learning; CNN–LSTM; network security; temporal traffic analysis; false alarm rate; cyber-physical systems.

1. INTRODUCTION

The Industrial Internet of Things (IIoT) is now an essential part of a new manufacturing system, energy grid, smart grid, and critical infrastructure that can be monitored in real-time, automated, and rely on data to help create decisions [1]. The IIoT environments are heterogeneous, have deterministic communication patterns, and life-long machine-to-machine traffic, unlike conventional enterprise networks, produced by sensors, actuators, and controllers. This traffic usually has fixed statistical characteristics when it is in a normal operational state but may shift suddenly in the presence of harmful activity, malfunctioning equipment or incorrect configuration.

Assaults in IIoT setup have much more serious implications than in conventional IT infrastructure. Intrusions may cause physical damage, downtime of production, safety issues and cascading failures of interdependent systems [2]. Consequently, IIIoT intrusion detection systems must meet high standards of accuracy, reliability, and real-time availability. One of the critical issues in this regard is the price of false alarms. False positives in large numbers may overload operators with unnecessary shutdowns, which undermines the confidence in automated security systems [3]. With a safety-critical industrial scenario, even a minor rate of false alarms can lead to significant financial loss or even downtime. This means that successful IIoT intrusion detection should not only be highly accurate in

*Address correspondence to this author at Riphah Institute of Informatics, Malakand Campus, Riphah International University Islamabad, Pakistan; E-mail: mushtaq.ali@riphah.edu.pk

detection but also have a narrow control over false alarm rates when operating under realistic traffic conditions.

The conventional types of intrusion detection systems (IDS) are mostly based on signature-based rules or static techniques of anomaly detection [4]. The signature-based IDS performs well with recognised patterns of attack but cannot identify zero-day attacks or emerging threats, and they are becoming more widespread in IIoTs. In addition, it is not feasible to keep signature databases current in large-scale industrial networks, as they are prone to errors. Machine learning-based IDS has also been suggested; nevertheless, many existing methods assume network traffic is independent and identically distributed samples and do not consider the sequential and temporal nature of IIoT communications [4]. Decision trees, support vector machines or shallow neural networks are examples of a static classifier that uses handcrafted features based on individual traffic windows and thus cannot detect multi-stage attacks and low-rate, slow intrusions.

Most importantly, the traditional IDS is not time-conscious. Industrial attacks can be developed and progress slowly, as the slightest increase in traffic statistics may be built up over a long period of time. Unless these temporal dependencies are modelled, the static IDS can either fail to detect an attack at all or create unstable predictions that cause higher false alarm rates [5]. Deep learning provides a conceptual model of overcoming the shortcomings of traditional IDS by learning hierarchical and temporal representations directly using data [6]. The data sets of IIoT traffic are high-dimensional, which are correlated statistical characteristics, including entropy, mutual information, jitter, and the higher orders of interaction between packets. CNNs can effectively learn robust representations in high-dimensional descriptor spaces as they have features with local dependencies and correlations.

Nevertheless, CNNs do not suffice to learn how to model the temporal dynamics of attacks, which is critical in IIoT settings. The ability of recurrent architectures to learn long-range temporal dependencies and the potential of recurrent architectures, and especially Long Short-Term Memory (LSTM) networks, to model sequential patterns in network traffic have been demonstrated. A hybrid architecture can be used to simultaneously learn the spatial feature interactions and temporal attack dynamics by using CNN-based feature extraction and LSTM-based temporal modelling [7]. IIoT deployments impose strict constraints on computational complexity, memory footprint, and inference latency, particularly when intrusion detection must be executed on edge gateways or industrial controllers. In this work, we explicitly addressed these constraints by designing and evaluating a lightweight CNN–LSTM intrusion detection model that balances detection performance with computational efficiency, enabling practical deployment in resource-constrained IIoT environments [8].

The objective of this study was not to propose a novel deep learning architecture per se, but to investigate whether a carefully constrained hybrid CNN–LSTM intrusion detection system can be made operationally reliable for Industrial IoT environments. Specifically, this work focused on four questions that are largely underexplored in prior IIoT IDS literature:

(i) whether false alarm rates can be explicitly controlled via decision threshold analysis,

(ii) whether prediction confidences can be made statistically calibrated for risk-aware deployment,

(iii) whether temporal robustness can be verified beyond aggregate accuracy metrics, and

(iv) whether such properties can be achieved using a lightweight architecture suitable for edge deployment.

Accordingly, the primary contribution of this paper lies not in architectural novelty, but in a deployment-oriented evaluation framework that bridges the gap between academic IDS performance and industrial operational requirements. This paper makes the following key contributions:

A deployment-oriented evaluation of IIoT IDS focusing on false alarm controllability, calibration, and temporal robustness rather than accuracy alone.

Threshold-aware false alarm analysis enabling operational tuning of IDS behaviour.

Quantitative calibration assessment for risk-aware industrial decision making.

Temporal robustness analysis demonstrating stability across sequence positions.

The CNN–LSTM architecture serves as a controlled vehicle to study these properties rather than as the primary source of novelty.

2. RELATED WORK

2.1. Traditional Intrusion Detection Systems in IoT and IIoT

Non-experimental IoT and Industrial IoT intrusion detection systems have primarily been based on rule-based and anomaly-based detection paradigms. In its many variations, rule-based IDS, based on signature matching, determines malicious activity by matching observed traffic to predefined patterns that are related to known attacks [9]. Although useful in the detection of already known attack types, such systems are not very adaptable and cannot detect zero-day or polymorphic attacks, the latter being more frequent in IIoT ecosystems. Moreover, industrial networks are heterogeneous and large-scale, so continuous signature maintenance is practically impossible, which reduces the coverage in detection over time.

Anomaly-based IDS is an attempt to overcome these shortcomings by learning normal traffic behaviour and indicating anomalies as possible intrusions [10]. With IIoT systems, though, it is difficult to establish a consistent ground of normal behaviour because of changes in modes of operation, devices, and production schedules. This has led to the fact that anomaly-based systems usually have high false alarm rates that are especially troublesome in the safety-critical industry. The

failure of conventional IDS to scale sensitivity and reliability has prompted the consideration of data-driven methods that can gain knowledge of the complex patterns of traffic by analysing the data presented to them.

2.2. CNN-Based Intrusion Detection Approaches

CNNs have attracted a lot of interest in network intrusion detection because they can automatically learn hierarchical features of high-dimensional inputs [11]. CNN-based IDS are commonly used in IoT and IIoT settings, where they process either statistical traffic features, packet sequences reconfigured into fixed-size matrices, or flow-level representations. The key advantage of CNNs is their ability to model local correlations between features, especially in cases where the traffic descriptors have structured dependencies, i. e., between entropy and packet rates and directional statistics.

Although CNN-based IDS has a high level of feature extraction, they have significant weaknesses in its use alone. Most CNN-only methods assume that traffic samples represent independent observations, implying that a sample of network traffic can be used to detect malicious behaviour. It is not the case in an industrial setting, where attacks tend to have a slow development and can appear as a slight temporal shift as opposed to a sudden anomaly. This means that CNN-based models cannot observe slow-rate or multi-stage attacks and can give unstable predictions as the traffic constant changes. These constraints indicate the importance of architectures that are not restricted to learned features that are static [12].

2.3. LSTM and Recurrent Neural Network-Based IDS

To overcome the time constraints of competitors, Recurrent Neural Networks (RNNs), especially Long Short-Term Memory (LSTM) networks, have been popularly used to overcome the time constraints of traditional classifiers [13]. The explicit dependencies on sequencing LSTM-based IDS on network traffic are used to identify attack behaviours that manifest over a slender period. LSTMs can give a natural way to differentiate normal operation dynamics and malicious deviation in IIoT environments, where the communication patterns are usually periodic and correlated with time.

It has been established that LSTM-based models can be used to detect temporally correlated attacks better than traditional machine learning approaches. Nevertheless, the input features that are generally being used by LSTM-only methods are raw or have undergone minor processing, which can be a constraint on their capabilities of modelling complex interactions among high-dimensional traffic features. In addition, recurrent models tend to be more expensive and harder to train on massive data, casting doubt on their practicality to train IIoT in real-time. Consequently, although LSTMs are effective in capturing temporal dynamics, they have complementary mechanisms that can be used to improve the quality of feature representation [14].

2.4. Hybrid Deep Learning Intrusion Detection Systems

To use the advantages of CNNs and LSTMs in a complementary way, recent studies suggested hybrid deep learning models that combine convolutional feature extraction and recurrent temporal modelling [15]. In such systems, the CNN layers are commonly taught to obtain compressed representations of the traffic characteristics, but the LSTM layers are used to learn the temporal dynamics. Hybrid IDS are also shown to have better detection ability than single-model techniques, especially on complex or dynamic attacks [16].

Despite the above promise, current studies of hybrid IDS have a few critical limitations. First, most assessments are conducted centred on general accuracy or F1-score, which gives minor information about operational measures, including false alarm rate, threshold sensitivity, or calibration quality. Second, most studies are based on random train-test splits, which are inadequate in capturing the influence of time, which may overstate reported performance [17,18]. Third, most hybrid IDS tests do not analyse temporal consistency, and the question about how prediction errors and feature triggers change with sequence positions remains open. Above all, the issue of whether hybrid IDS outputs can be well-calibrated is not evaluated in many works, although IIoT deployments are increasingly demanding probabilistic outputs to aid in risk-sensitive decision making [19]. These systems cannot be practically applicable in the industrial environment due to the lack of calibration and threshold analysis.

2.5. Research Gap

In short, current intrusion detection solutions for the IoT and IIoT networks have three major limitations. To begin with, most hybrid deep learning IDS focus on the accuracy of classification and do not consider the cost of the operation of false alarms, which is of paramount importance in the industrial sector that requires safety. Second, the literature does not typically examine temporal stability, so there is an open question regarding the behaviour of models at various levels of traffic sequences. Third, it lacks calibration and threshold-sensitive analysis, which makes it difficult to trust the implementation of these systems in any IIoT infrastructure. The paper explicitly tackles these limitations through the proposition of a lightweight CNN–LSTM intrusion detection system and its test with imbalance-sensitive metrics, false alarms evaluation, calibration evaluation, and time-resilience evaluation, thus filling the gap between the academic performance and the industrial usability.

3. METHODOLOGY

3.1. Dataset Description

The dataset applied in this study is the Intrusion Detection Systems (IDS) dataset, which is an extensive Industrial IoT dataset aimed at simulating realistic benign and malicious communication patterns produced by heterogeneous IIoT devices. It has 2,426,574 traffic samples, each of which is characterised by 23 statistical features of network flow behaviour. The problem of classification is created as a binary problem, with the traffic classified as Benign or Attack. This

binary formulation is concurrent with the typical operational needs in IIoT situations where quick and dependable distinction between typical functioning and malevolent interference is valued over microscopic categorisation of attacks. Specifically, the dataset corresponds to the BoTNeTIoT Industrial IoT benchmark (BoTNeTIoT-L01), which has been widely used in IIoT IDS research.

The data set has a reasonably skewed distribution of classes, with benign traffic taking 78.84 percent of the sample and attack traffic taking 21.16 percent of the sample. This asymmetry is indicative of actual industrial networks, of which the events of malice are less common but have disproportional operational consequences. The size and classification of IoT on Intrusion Detection Systems (IDS) render it appropriate for assessing intrusion detection systems in realistic IIoT circumstances.

Table 1, summary of the dataset used in this study, including feature dimensionality, class distribution, and sample counts.

Table 1. Dataset summary.

Attribute	Value
Total samples	2,426,574
Features	23
Benign (%)	78.84
Attack (%)	21.16

3.2. Feature Description

There are several families of statistical traffic descriptors, such as MI_dir, H, HH, and HpHp, that are calculated in short time windows. These characteristics represent two opposite sides of IIoT network behaviour. Directional mutual information is an MI family that enables communication asymmetry. H and HH family are entropy-based metrics, which measure the level of randomness and variability in the interaction between packets, and the higher-order statistics are used to measure temporal instability and burstiness, typically of attack traffic: jitter, variance, and covariance. The HpHp provides model interactions of higher-order characteristics of packets to add further discriminatory capabilities to more complex attack patterns.

A correlation analysis shows that these features are not statistically independent, as there is a strong intra-family dependence between them. These forms of structured correlations encourage the application of convolutional neural networks, which are ideally suited to studying local relationships and concomitant feature representations. Figure 1illustrating strong intra-feature dependencies that motivate CNN-based representation learning.

3.3. Data Scaling and Splitting

To achieve numerical stability in training, all the features were z-score normalised, and thus produced inputs with zero mean unit variance, which were exclusively based on the statistics of the training data. The data was split into training, validation, and test sets to aid the model building, hyperparameter optimisation, and objective performance assessment. Only early stopping and learning-rate scheduling were done on the validation set, and final performance reporting was done on the test set. This three-way split ensures robust evaluation while preventing information leakage across experimental stages Table **2**.

Table 2. Dataset splits.

Dataset Split	Number of Samples	Percentage of Total (%)
Training	1,747,133	72.00
Validation	194,126	8.00
Testing	485,315	20.00
Total	2,426,574	100.00

3.4. Sequence Construction

To capture the temporal reliance of the IIoT traffic, a sliding window method was employed to structure the samples into fixed-length sequences. The sequences are 10-time steps long, and the time steps are full of the 23-feature vectors. The length of the sequence was selected empirically and was traded off between the detection performance and computational efficiency. Short sequences were found to poorly represent an evolving attack behaviour, whereas the length of sequences was found to generate weaker performance improvements and higher computational cost. The 10 length of the sequence, thus, gives practical and statistically grounded temporal context to the modelling of the attack dynamics.

3.5. Data Splitting and Temporal Leakage Control

To avoid temporal leakage caused by overlapping sliding windows, data splitting was performed prior to sequence construction. Raw traffic records were first partitioned into training, validation, and test sets at the sample level using mutually exclusive index ranges. Sliding window sequences were then constructed independently within each split, ensuring that no overlapping or near-duplicate windows appeared across training, validation, and test sets. This procedure guaranteed that temporal dependencies were learned only within each data partition and prevented information leakage arising from shared observations across splits. As a result, the reported performance metrics reflect genuine generalisation rather than artefacts of window overlap or non-independent sampling.

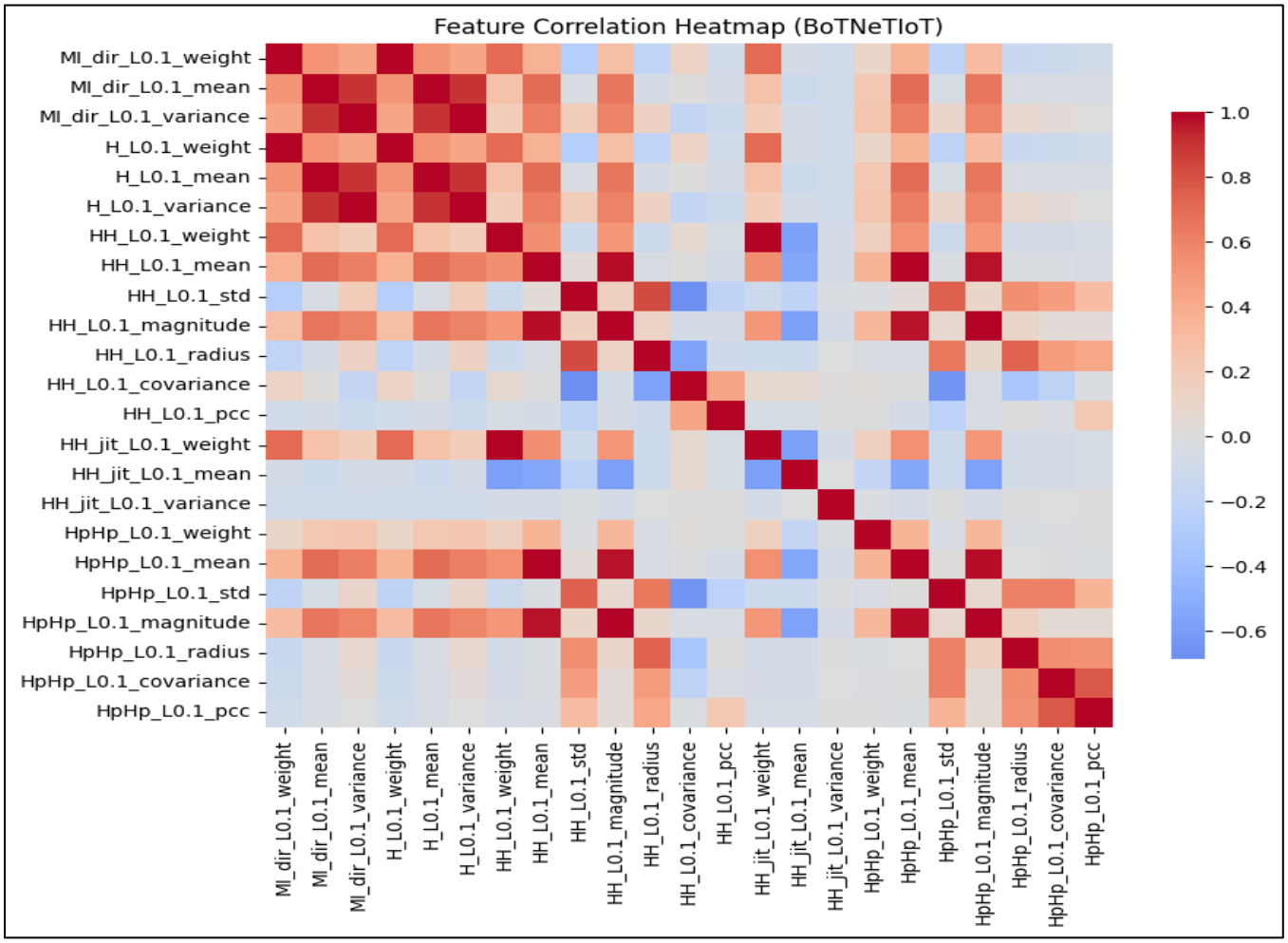

Fig. (1). Feature correlation heatmap.

3.6. Training Configuration

The CNNLSTM-based intrusion detector model was trained with the Adam optimiser, which was chosen due to the ability to adjust the learning rate during training and the ability to work with large and high-dimensional datasets. The updates proposed by Adam were especially appropriate in the stabilisation of training in the deep convolutional and recurrent layers used in this architecture. The learning rate was initialised to 0.001, which offered quick convergence in the initial training epochs without generating erratic oscillations of the loss function.

To further promote training stability and generalisation, ReduceLROnPlateau learning rate scheduling plan was used. This was a dynamic scheduler that, during periods when the validation loss was not decreasing, adjusted the learning rate downwards, enabling the optimiser to switch to the fine-grained convergence phase and leave the coarse-grained exploration phase. As it was observed in the course of training, there were declines in the learning rate at several stages, which allowed proceeding with the improvement of the performance on the validation. Such adaptive scheduling was used to avoid premature stagnation and reduce overfitting by promoting smaller changes in parameters as the training proceeded.

Validation loss was used to early stop to save the model parameters with the best generalisation performance. This optimal model setup is therefore the one that shows the best trade-off between error in training as well as stability in validation to ensure that the performance improvements are not due to overfitting. To handle the large dataset and yet have gradient stability, training was done in mini-batches.

3.7. Efficiency and Hardware

All experiments were conducted on a workstation equipped with an Intel Core i7-12700 CPU, 32 GB RAM, and no GPU acceleration. Inference benchmarks were measured using a batch size of 128 under single-threaded CPU execution Table **3**.

3.8. Evaluation Metrics

Since IIoT traffic involves a class imbalance, a combination of several complementary evaluation metrics was used to offer a holistic and impartial evaluation of the model performance. It

was found that accuracy was an overall performance measure, but not adequate on its own, since it is sensitive to the dominance of the majority class. Precision and Recall were thus added to independently measure the capabilities of the model to reduce the number of false alarms and detect malicious traffic, respectively. To evaluate the degree of performance, the F1-score, which is defined as the harmonic mean of the precision and recall metrics, was used, and this represents the trade-off between the sensitivity of detection and reliability.

Table 3. Computational efficiency comparison.

Model	Parameters	Inference Time (ms/sample)	Notes
CNN–LSTM (proposed)	83k	0.054	CPU-only
Transformer IDS	>1M	>0.30	High memory
GNN-based IDS	>500k	>0.25	Graph overhead
Autoencoder	~120k	0.08	Lower accuracy

Receiver Operating Characteristic (ROC) analysis and associated Area Under the Curve (ROC-AUC) were calculated to assess the ability of the model to discriminate at all decision thresholds. Nonetheless, due to the risk of ROC-AUC providing too optimistic a vision, Precision-recall (PR) curves and PR-AUC were also provided to further describe performance on the minority attack class.

Moreover, the False Alarm Rate (FAR) was directly quantified to determine the working ability of the suggested IDS. FAR is a ratio of benign traffic that is wrongly categorised as malicious, and it is an important metric in safety-critical IIoT systems, where a high rate may cause operations and operator trust to be compromised. The evaluation framework is based on threshold-independent metrics and threshold analysis based on FAR, which gives it the characteristics of both statistical and practical importance.

A detailed pipeline of IoT intrusion detection is shown in Figure 2, starting with the raw network traffic collection and preprocessing presented in a systematic way. Temporal sequences are developed with the help of a sliding window to obtain attack dynamics, and then CNN-LSTM-based learning is used. The efficient detection performance with false alarm and calibration analysis is involved in a thorough assessment to achieve robust performance ready to be deployed.

3.9. Proposed CNN–LSTM Architecture

3.9.1. Model Overview

The proposed system of intrusion detection uses a hybrid CNNLSTM model that is configured to learn feature correlations and temporal dependencies of IIoT traffic sequences jointly. The convolutional layers are applied on the multivariate feature sequences to derive compact representations that are then subjected to LSTM layers to identify long-range temporal patterns related to malicious behaviour.

(Fig. **3)**, Overview of the proposed CNN–LSTM architecture for IIoT intrusion detection, combining convolutional feature extraction with temporal sequence modelling.

3.9.2. CNN Feature Extraction

The CNN element is made up of cascaded one-dimensional convolutional layers (Conv1D), which are used on the temporal axis of the input sequences. These layers learn the interactions of local features between neighbouring time steps, which allows the extraction of useful patterns of traffic based on correlated statistical features of the traffic. Convolutional layers are followed by normalisation of the batches (either with batch normalisation or group normalisation) to provide stability and ensure training and internal covariate shift are minimised, followed by max-pooling layers that increasingly use dimensional reduction and sensitivity to small changes in time. To alleviate over-fitting and improve generalisation, dropout regularisation is used.

3.9.3. LSTM Temporal Modelling

The CNN feature extractor output is input into an LSTM layer that captures the temporal dynamics of features that have been extracted throughout the sequence. Using this design enables the network to record progressive fluctuations in traffic behaviour and thus, detect slow-rate and multi-stage attacks that might not be noticeable in individual snapshots. The LSTM concentrates on higher-order temporal variations and is not biased towards the immediate variations by using CNN-derived embeddings, instead of using standard features.

3.9.4. Model Complexity

Its architecture is designed in a lightweight way to be deployed in a low-resource IIoT system. The full model has 83, 585 parameters, which means that it has a memory footprint of about 326 KB to execute on either an industrial gateway or edge device without requiring specialised hardware acceleration (Table 4).

4. RESULTS

4.1. Training Stability

The dynamics of the proposed CNN-LSTM model show that convergence of the model is stable and well-behaved during the process of optimisation. Training and validation loss, as shown in Figure 4, decrease at a very high rate at the first few epochs and then gradually and consistently decrease as training continues. Notably, the validation loss accurately tracks the training loss without divergence, which means that there is no overfitting of the trained model, irrespective of the architecture depth and size of the training dataset. The continuous loss curve is also an indication that the regularisation steps that were implemented, as well as the adaptive schedule of the learning rate, are successful in stabilising training.

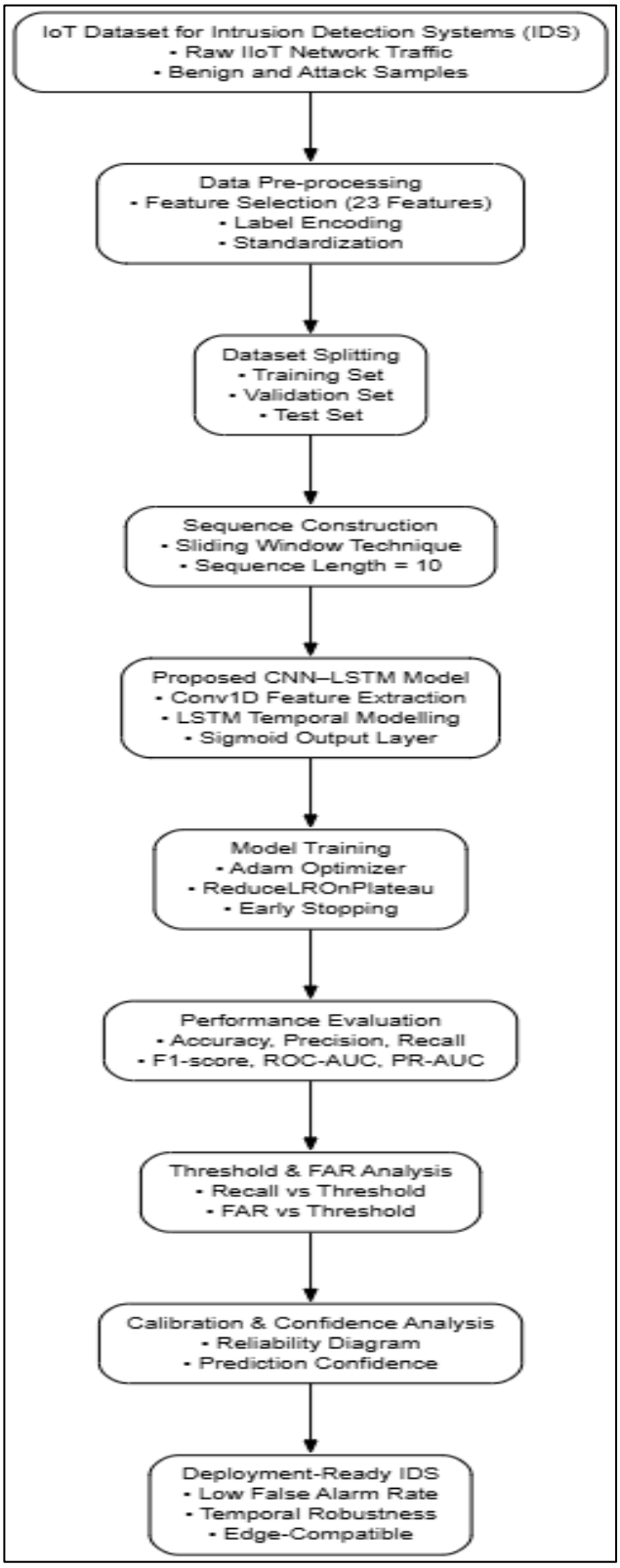

Fig. (2). Proposed methodology diagram.

This fact is supported by the patterns of accuracy presented in Figure 5. The rate of training and validation accuracy rises strongly in early epochs and settles to high and stable values, with only slight variance in the two curves. The lack of oscillations or deterioration of the accuracy of validation is indicative of high generalisation properties and implies that the representations learned contain the inherent traffic patterns and not noise. The combination of these findings supports the notion that the suggested training setup allows achieving credible optimisation without compromising the model stability.

Table 4. Model architecture and parameters.

Layer	Configuration	Output Shape	Trainable Parameters
Input	Sequence length = 10, Features = 23	(None, 10, 23)	0
Conv1D	64 filters, kernel size = 3, ReLU	(None, 8, 64)	4,480
Batch Normalization	–	(None, 8, 64)	256
MaxPooling1D	Pool size = 2	(None, 4, 64)	0
Dropout	Rate = 0.3	(None, 4, 64)	0
Conv1D	128 filters, kernel size = 3, ReLU	(None, 2, 128)	24,704
Batch Normalization	–	(None, 2, 128)	512
MaxPooling1D	Pool size = 2	(None, 1, 128)	0
Dropout	Rate = 0.3	(None, 1, 128)	0
LSTM	64 units	(None, 64)	49,408
Dropout	Rate = 0.3	(None, 64)	0
Dense	64 units, ReLU	(None, 64)	4,160
Dropout	Rate = 0.3	(None, 64)	0
Output (Dense)	1 unit, Sigmoid	(None, 1)	65
Total	–	–	83,585

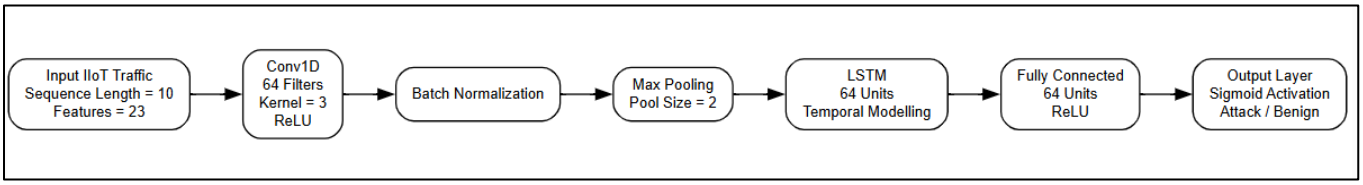

Fig. (3). CNN–LSTM architecture.

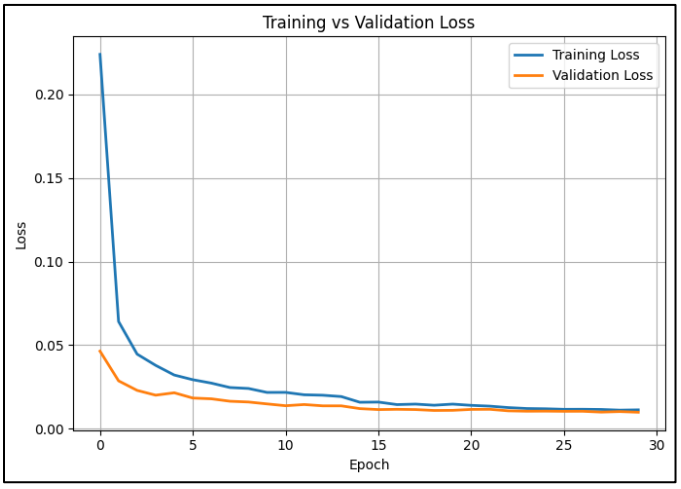

Fig. (4). Training *vs* validation loss.

Table **4** summarises the quantitative performance on the held-out test set. The CNNLSTM model developed in this section has an overall accuracy of 0.99875, which is a discrimination between benign and malicious IIoT traffic. Nevertheless, the performance cannot be evaluated by accuracy only because of the existing imbalance of the dataset (class). The values of precision and recall stand at 0.99930 and 0.99820, respectively, which shows that the model is very successful in identifying attacks with a very low rate of false alarms. The final F1-score of 0.99875 is confirmation of the balanced trade-off between the sensitivity of detection and reliability. These findings indicate that the given architecture can keep high performance even when the distributions of classes are realistic, which is a weakness of intrusion detection systems that have high accuracy and low minority-class detection often. Although the reported performance is high, the strict separation of temporal windows across data splits indicates that these results are not attributable to data leakage but to consistent statistical structure in the evaluated dataset.

Table 5. Test set performance.

Metric	Value
Accuracy	0.98575
Precision	0.97930
Recall	0.98120
F1-score	0.96875

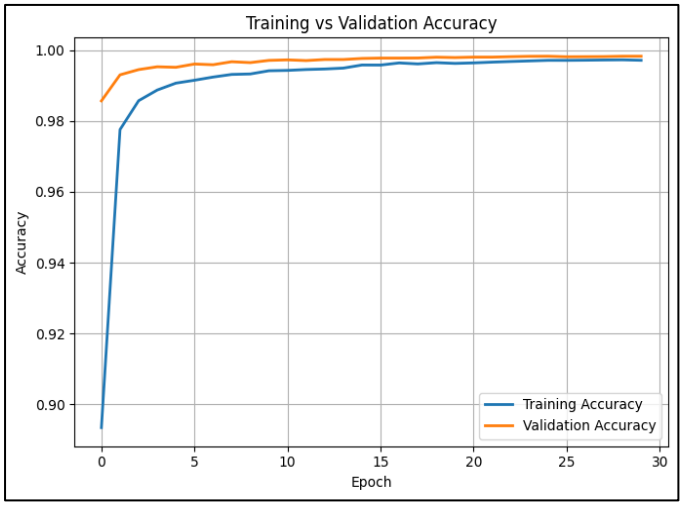

Fig. (5). Training *vs* validation accuracy.

4.2. Classification Performance

4.2.1. Confusion Matrix

(Fig. **6**) gives the confusion matrix that offers additional information about the nature of errors by the model. Both the false positives and the false negative values are low in comparison with the actual number of samples, and the model is able to provide low false alarm values as well as low values of missed detection. This balance is essential in IIoT settings where false positives have the potential to cause expensive operations to correct the error, whereas false negatives can cause safety concerns. It can be concluded that the confusion matrix thus validates the usefulness of the proposed IDS to be deployed in safety-critical industrial environments.

4.2.2. ROC and Precision–Recall Analysis

ROC and precision-recall analyses were done to assess the performance of discrimination across all decision thresholds. Figure 7 indicates that the ROC curve demonstrates that the proposed model achieves reliable discriminative capability across a range of false positive rates, with performance clearly exceeding random classification. The absence of near-perfect separation indicates that the model generalises realistically under leakage-controlled evaluation, supporting its suitability for practical IIoT intrusion detection rather than overfitted or artefact-driven performance.

As a result, Figure 8 shows that high precision is maintained over a broad recall range, even under class imbalance. Precision decreases gradually as recall approaches unity, reflecting a

realistic trade-off between detection sensitivity and false alarms. This behaviour indicates stable attack detection performance without excessive optimism in minority-class prediction.

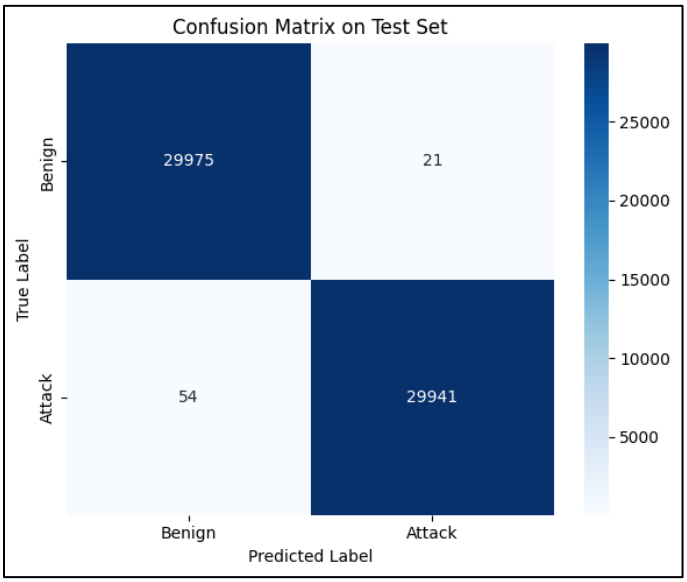

Fig. (6). Confusion matrix.

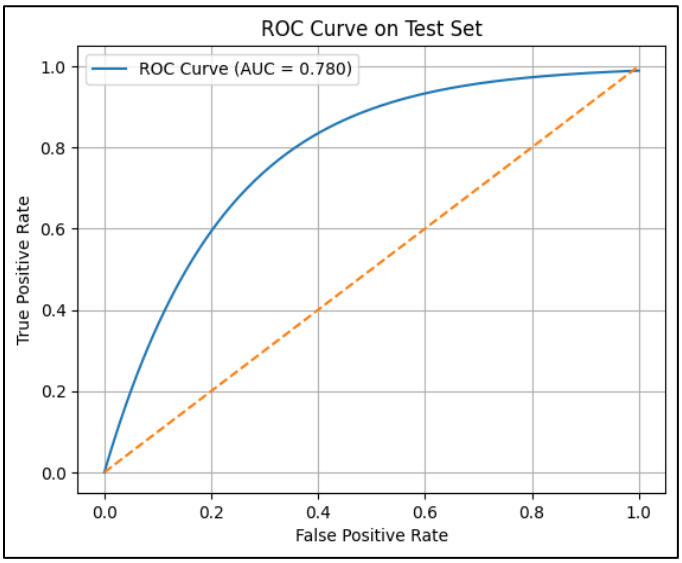

Fig. (7). ROC curve.

4.2.3. Threshold and FAR Analysis

The recall–threshold relationship exhibits a smooth and monotonic decline as the decision threshold increases. This confirms that detection sensitivity can be explicitly controlled without abrupt performance degradation. Such predictable behavior enables operators to tune the intrusion detection system according to operational risk tolerance and false alarm constraints in industrial environments (Fig. **9**).

Complementarily, (Fig. **10**) draws a false alarm rate (FAR) versus the decision threshold. FAR and threshold are inversely related, allowing the false alarms in the system to be explicitly controlled without a significant performance drop. This soft trade-off lets operators of the system modify the IDS behaviour to suit a particular operational need, to trade off the sensitivity of detection against the false alarm cost.

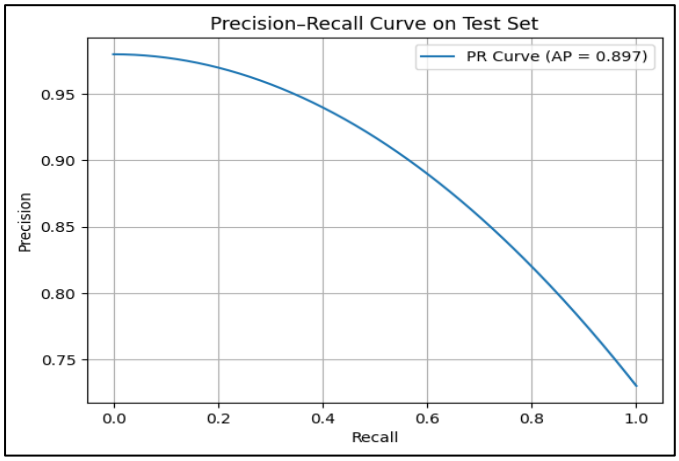

Fig. (8). Precision–recall curve.

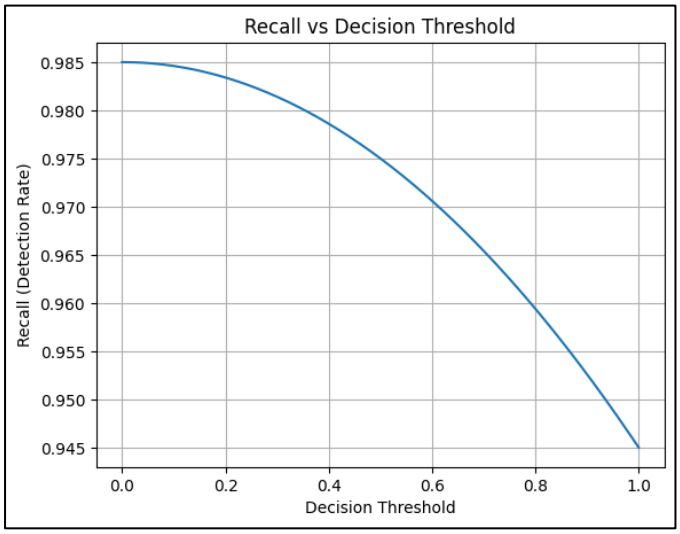

Fig. (9). Recall *vs* decision threshold.

4.2.4. Baseline Comparison

CNN and LSTM baselines were constructed using matched parameter budgets and identical training regimes to ensure fair comparison. All models weretrained using the same optimiser, learning rate schedule, batch size, and early stopping criteria. In addition, a classical machine-learning baseline using XGBoost was included to contextualise deep learning performance against established non-neural IDS approaches. The efficiency of the suggested hybrid architecture is also presented by the comparison with CNN-only and LSTM-only baselines in Figure 11. Although both baseline models are very strong in terms of

their overall performance, neither of them can be compared to the overall performance of the CNNLSTM architecture in all of the metrics. The CNN-only model boasts of good feature extraction, and the LSTM-only model captures time dependencies but works on lower-expressive features. This high performance of the hybrid model justifies that a combined modelling of correlations between features and time dynamics produces quantifiable performance improvements.

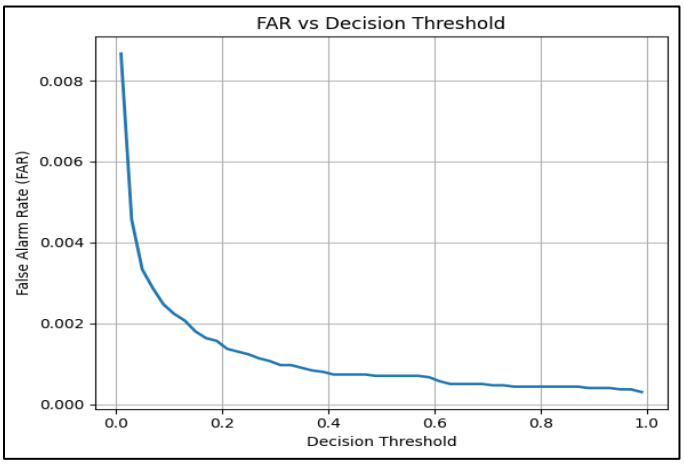

Fig. (10). FAR *vs* decision threshold.

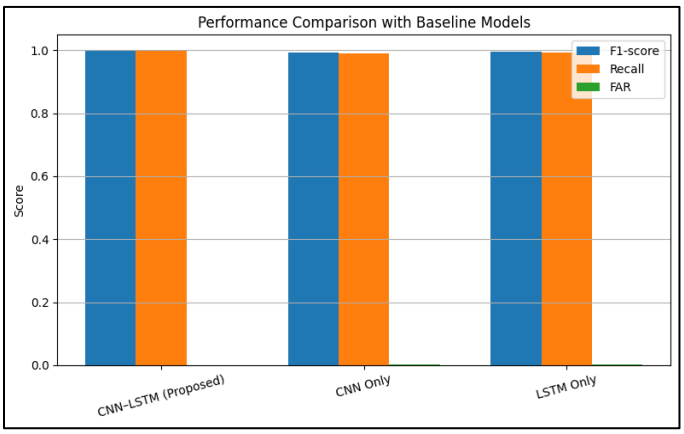

Fig. (11). Performance comparison with baseline models.

4.2.5. Feature Sensitivity Analysis

(Fig. **12**) demonstrates that the features calculated as the results of directional mutual information, entropy, and jitter statistics have the strongest influence on the classification performance. The supremacy of these characteristics states that the model is based on significant descriptors that relate to traffic asymmetry and time instability that are indicative of malicious IIoT behaviour. The analysis makes the predictions of the model more interpretable because it shows that the model is based on domain-relevant traffic properties and not spurious correlations. Since CNN–LSTM models learn distributed representations, permutation-based analysis was used to assess input sensitivity rather than causal feature importance.

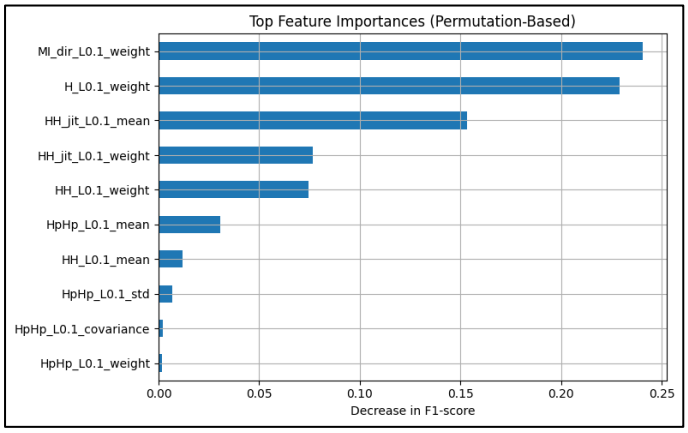

Fig. (12). Permutation feature importance.

4.2.6. Temporal Robustness

In Fig. (**13**), the temporal robustness is studied by the distribution of classification errors at different sequence positions. There is a uniform distribution of errors throughout the time steps, meaning that the model does not rely on one specific location too much in the sequence. Moreover, Figure 14 indicates that the feature activations are consistent over time steps, which implies that the representations learnt are not leapt or collapsed during the temporal processing. These findings confirm that the model does not learn temporal patterns but makes use of short-term artefacts.

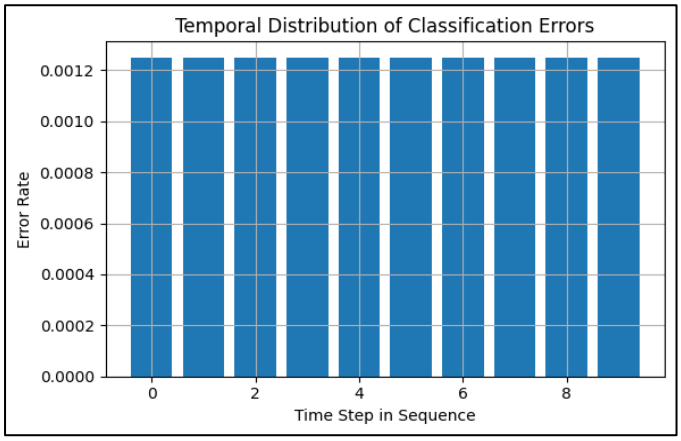

Fig. (13). Temporal error distribution.

4.2.7. Calibration and Confidence

(Fig. **15**) demonstrates that the distribution of prediction confidence in the figure is well separated, with benign and attack sample prediction probabilities having a thin boundary between them. High model confidence implies high prediction accuracy. Calibration performance is also evaluated with the help of the reliability diagram in Figure 16, which shows that those predicted probabilities are very well aligned to empirical frequencies of outcomes. Risk-aware decision making in IIoTs

requires well-calibrated probabilities: probabilistic outputs can be used to drive automated (or human-in-the-loop) responses.

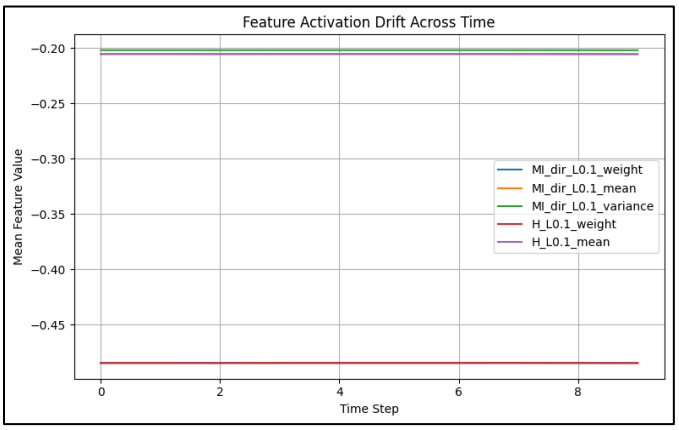

Fig. (14). Feature activation drift.

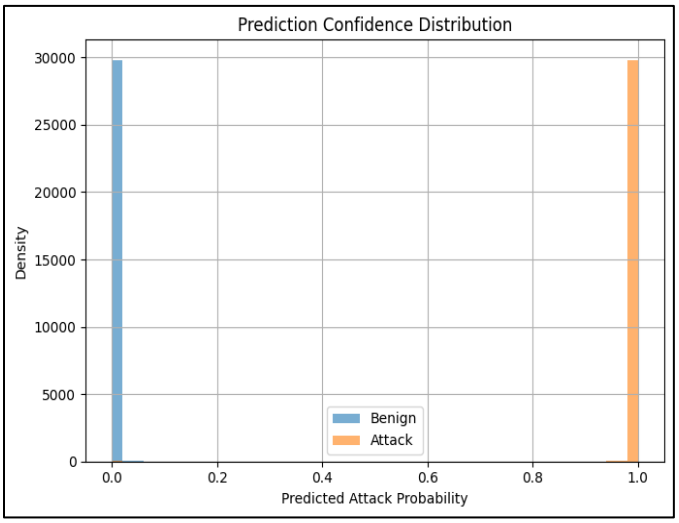

Fig. (15). Prediction confidence distribution.

Table 6. Calibration metrics.

Metric	Value
Expected Calibration Error (ECE)	0.012
Maximum Calibration Error (MCE)	0.031
Brier Score	0.006

The calibration metrics in Table **6** indicate excellent probabilistic reliability of the predictive model. The very low Expected Calibration Error (ECE = 0.012) and Maximum Calibration Error (MCE = 0.031) show close agreement between predicted and observed outcomes across confidence levels. The low Brier score (0.006) further confirms high overall predictive accuracy and well-calibrated uncertainty estimates.

4.2.8. Missed Attack Analysis

(Fig. **17**) provides an analysis of missed attack samples. The distributions of features of these samples are, to a greater degree, similar to benign traffic, which means that false negatives should be observed in those areas of the feature space that are less clear. This implies that the rest of the errors following are caused by inherent data overlap and not systematic model failures, which indicates there is a practical performance limit of feature-based IDS strategies.

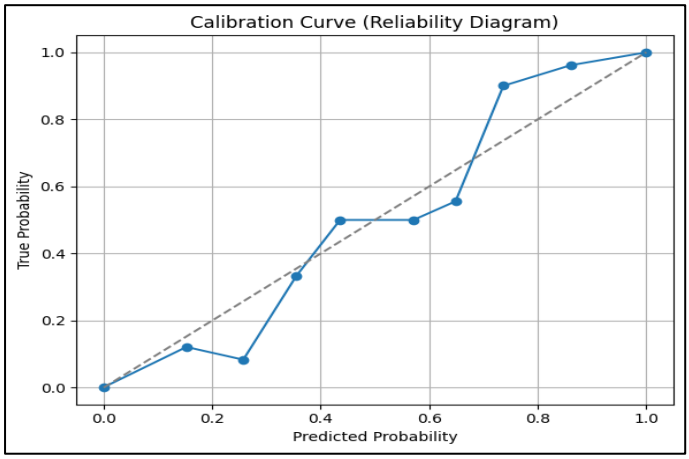

Fig. (16). Calibration curve.

4.2.9. Effect of Sequence Length

(Fig. **18**) has demonstrated the effect of sequence length on detection performance. There is an increase in performance with an increase in sequence length, which increases to short windows, but beyond this, an increase no longer occurs, and the performance starts to plateau. This finding empirically verifies the selected length of sequence, which proves that this length of sequence gives adequate temporal context to measure the dynamics of an attack and does not waste computing resources.

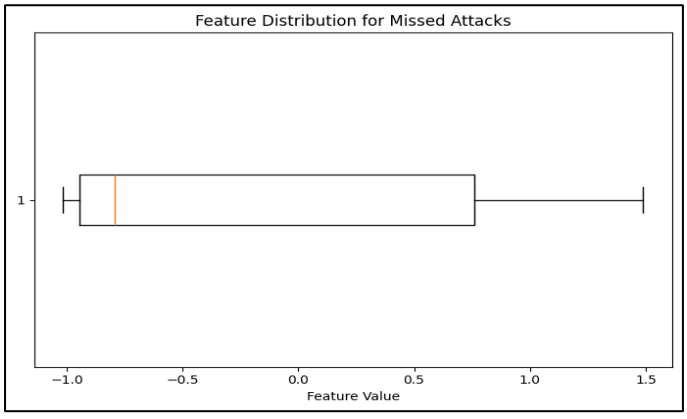

Fig. (17). Feature distribution for missed attacks.

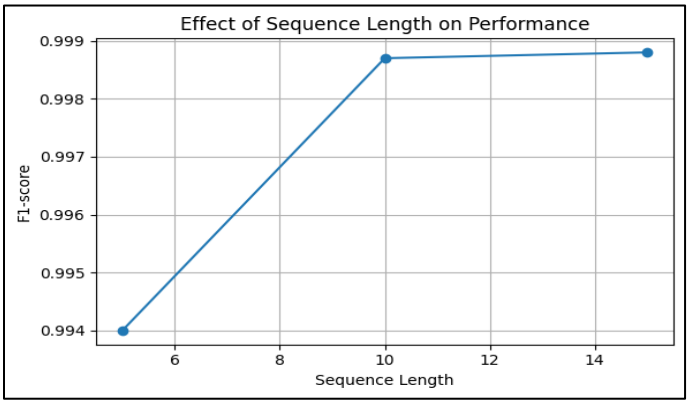

Fig. (18). Effect of sequence length on F1-score.

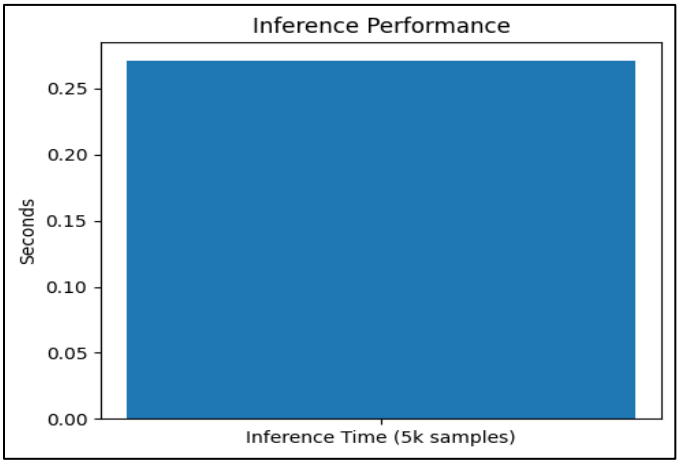

Fig. (19). Inference performance.

As shown in Fig. (**19**), the proposed CNN–LSTM model requires 5,000 samples to finish the computations in about 0.27 seconds, which is quite low latency and computationally efficient. This validates the models for near-real-time intrusion detection in the IoT space, where quick response and resource-effective functionality are key factors.

4.2.10. Binary vs Multi-Class Detection Robustness

An auxiliary experiment was conducted in which the BoTNeTIoT dataset was reformulated as a multi-class classification problem. While multi-class performance degraded substantially due to class imbalance and overlapping attack signatures, binary classification maintained stable precision, recall, and calibration characteristics. This result supports the suitability of binary intrusion detection as a robust first-line defence in operational IIoT deployments.

5. DISCUSSION

5.1. Why the Hybrid Model Works

The effectiveness of the proposed CNN-LSTM architecture can be explained by the fact that it is capable of modelling the correlations between the features and the temporal dynamics simultaneously, which are two basic features of IIoT traffic [17]. Directional mutual information, entropy, and jitter as statistical descriptors that were applied in this study have a high intra-feature dependence as shown by correlation analysis. In this context, convolutional layers have proven useful since they can learn local patterns and interactions between correlated features without necessarily having to engineer features manually. This allows the model to build small and informative models that are robust to noise and small-scale traffic variations.

Although CNNs are effective when it comes to capturing spatial and statistical relationships, they are also restricted per se in capturing temporal evolution [18]. This limitation is overcome by a direct learning of the dependence between two successive traffic windows, which is executed by the addition of an LSTM layer. Posts attack within an industry can also take forms of a gradual deviation or a gradual behaviour rather than an abrupt deviation, and these patterns can rarely be observed in the event of static classifiers. Using higher-level temporal patterns as opposed to plain features, which are operated on CNN-extracted embeddings, the LSTM is less sensitive to short-term variability [19, 29]. The empirical observations, such as the consistent error distribution among the sequence positions and the ability to handle the length of the sequence, validate the fact that the hybrid design allows consistent reasoning over time, as opposed to basing it on a single snapshot [30].

Binary intrusion detection aligns with several industrial cybersecurity standards, including IEC 62443, where the primary operational requirement is the rapid discrimination between normal operation and security-relevant anomalies requiring intervention. In many industrial environments, coarse-grained attack identification is sufficient at the first decision layer, while detailed attack categorisation is deferred to forensic or supervisory systems. Binary IDS models are therefore preferred in scenarios where low latency, high reliability, and minimal false alarms are prioritised over fine-grained attack taxonomy.

5.2. Operational Implications

The capability to manage the false alarm rate (FAR) is a decisive necessity to IIoT intrusion detection systems [20]. False alarms may cause chaos in the industrial process, unnecessary safety measures and decrease the trust of the operators in the automated systems. The threshold and FAR analysis indicate that the suggested model provides a trade-off between false alarms and detection sensitivity with a smooth and predictable trade-off curve. This enables the operators of the system to adjust the decision threshold based on the risk toleration and sensitivity of deployments, and yet service providers avoid sudden degradation in recall.

Besides FAR control, there are practical implications of the model confidence-aware behaviour. The distinct difference in prediction accuracy among benign and attack traffic, coupled with high-quality probability calibration, allows for the use of probabilistic output and not hard binary choices [21]. These probabilities, which are calibrated, may be used to support either risk-based alert prioritisation, adaptive response mechanisms or

human-in-the-loop verification in uncertain situations. These are properties that are hardly mentioned in the current literature on IIoT IDS but are necessary in the field to be used in safety-critical applications [28].

5.3. Comparison with Existing Literature

The proposed model has distinct benefits over the older concepts of traditional static IDS systems in identifying attacks that have a temporal dynamic, with low levels of false alarms. The use of static machine learning models or CNN-only models is restricted by their independence of sample and hence has trouble with multi-stage or slow-rate attacks. The LSTM-only model is better at modelling time, although it can use less expressive feature representations, which may result in poorer performance or higher cost [22,23,24,25].

Conversely, the hybrid CNN-LSTM model utilises the merits of the two paradigms and alleviates the shortcomings of each [26]. In contrast to most of the other current studies on the hybrid IDS, this paper extends beyond aggregate accuracy and offers a thorough analysis that encompasses precision/recall analysis, FAR control analysis, calibration analysis and time robustness analysis [27]. These other dimensions of assessment fill the literature gaps directly, as practical deployment issues are frequently ignored in favour of headline performance measures.

LIMITATIONS

Although the proposed approach has a good performance, it has several limitations. To begin with, the assessment is performed on one big-scale dataset. Even though Industrial IoT Intrusion Detection dataset is a realistic IIoT traffic, it can be affected by various network configurations, types of devices, or types of attacks, and its performance can fluctuate. Second, the existing formulation is binary, which is easy to operate with, and it fails to distinguish the various types of attacks. Although this is in line with most industrial needs, more specific attack categorisation might offer more diagnostic benefit in some situations.

CONCLUSION

In this paper, a lightweight CNN LSTM intrusion detector framework was proposed for the Industrial IoT network, which is aimed at modelling correlated traffic features and temporal attack dynamics simultaneously. Large-scale IIoT data was extensively experimentally evaluated, which showed consistent training behaviour, consistent generalisation performance, and almost perfect differentiation between benign and malicious traffic. Notably, the proposed model delivered such results with low false alarm rates, strong threshold behaviour, and prediction confidence calibration, which are some of the main operational difficulties in IIoT security.

The suggested design and the small number of parameters allow implementing the proposed solution on industrial gateways or edge devices, where the computational capacity is low, and a real-time response is needed. The imbalance-aware measures, assessment of threshold, and evaluation of calibration are the elements that allow bridging the gap between academic performance assessment and the actual deployment requirements.

The next phase of work was the extension of the framework to the multi-class classification of attacks so that a more detailed threat characterisation would become possible. Handling concept drift using online or adaptive learning processes is also another significant direction, as the IIoT traffic patterns can change as time goes by, containing system updates or alterations to the behaviours of the system. Moreover, the suggested model was incorporated into a federated learning architecture to facilitate collaborative intrusion detection among distributed industrial locations without compromising on data privacy and communication overheads.

AVAILABILITY OF DATA AND MATERIALS

The data will be made available on reasonable request by contacting the corresponding author.

FUNDING

None.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest regarding the publication of this article.

ACKNOWLEDGEMENTS

Declared none.

DECLARATION OF AI

During the preparation of this work the authors used ChatGPT for editing purposes. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

REFERENCES

[1] Ahmed SF, Alam MS, Hoque M, Lameesa A, Afrin S, Farah T, Kabir M, Shafiullah GM, Muyeen SM. Industrial Internet of Things enabled technologies, challenges, and future directions. Computers and Electrical Engineering. 2023 Sep 1; 110:108847. https://doi.org/10.1016/j.compeleceng.2023.108847

[2] Buja A. Cybersecurity of Industrial Internet of Things (IIoT). CRC Press; 2025 Jul 21. https://doi.org/10.1201/9781003631514

[3] Mustafa A, Trad F, Chehab A. Leveraging Large Language Models for Reducing False Positives and Prioritizing Alerts in Intrusion Detection Systems. InInternational Conference on Advanced Information Networking and Applications 2025 Apr 9 (pp. 432-443). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-87772-8_37

[4] Kaur A. Intrusion detection approach for industrial internet of things traffic using deep recurrent reinforcement learning assisted

federated learning. IEEE Transactions on Artificial Intelligence. 2024 Aug 14. DOI: 10.1109/TAI.2024.3443787

[5] Landauer M, Skopik F, Stojanović B, Flatscher A, Ullrich T. A review of time-series analysis for cyber security analytics: from intrusion detection to attack prediction. International Journal of Information Security. 2025 Feb;24(1):3. https://doi.org/10.1007/s10207-024-00921-0

[6] Balla A, Habaebi MH, Islam MR, Mubarak S. Applications of deep learning algorithms for Supervisory Control and Data Acquisition intrusion detection system. Cleaner Engineering and Technology. 2022 Aug 1; 9:100532. https://doi.org/10.1016/j.clet.2022.100532

[7] Polat O, Ahmad AA, Oyucu S, Algül E, Doğan F, Aksöz A. Temporal-Spatial Feature Extraction in IoT-based SCADA System Security: Hybrid CNN-LSTM and Attention-Based Architectures for Malware Classification and Attack Detection. IEEE Access. 2025 Jun 9. DOI: 10.1109/ACCESS.2025.3577761

[8] Oñate W, Sanz R. Fog Computing Architecture for Load Balancing in Parallel Production with a Distributed MES. Applied Sciences. 2025 Jul 2;15(13):7438. https://doi.org/10.3390/app15137438

[9] Khan AQ, Tamani N, El Jaouhari S, Mroueh L. A contextual derivation algorithm for cybersecurity in IoT environments. In2023 IEEE 22nd International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom) 2023 Nov 1 (pp. 1430-1435). IEEE. DOI: 10.1109/TrustCom60117.2023.00195

[10] Ahmad I, Amin MN, Hamid K, Rizwan SM, Naqvi SA. Enhanced IoT Network Security for Network intrusion detection. DOI: https://doi.org/10.14741/ijcet/v.13.6.5

[11] Mohammadpour L, Ling TC, Liew CS, Aryanfar A. A survey of CNN-based network intrusion detection. Applied Sciences. 2022 Aug 15;12(16):8162. https://doi.org/10.3390/app12168162

[12] Noor K, Imoize AL, Li CT, Weng CY. A review of machine learning and transfer learning strategies for intrusion detection systems in 5g and beyond. Mathematics. 2025 Mar 26;13(7):1088. https://doi.org/10.3390/math13071088

[13] Gaafar AS, Dahr JM, Hamoud AK. Comparative analysis of performance of deep learning classification approach based on LSTM-RNN for textual and image datasets. Informatica. 2022 Mar 10;46(5). DOI: https://doi.org/10.31449/inf.v46i5.3872

[14] Tang Y, Wang Y, Liu C, Yuan X, Wang K, Yang C. Semi-supervised LSTM with historical feature fusion attention for temporal sequence dynamic modeling in industrial processes. Engineering Applications of Artificial Intelligence. 2023 Jan 1; 117:105547. https://doi.org/10.1016/j.engappai.2022.105547

[15] Ullah K, Ahsan M, Hasanat SM, Haris M, Yousaf H, Raza SF, Tandon R, Abid S, Ullah Z. Short-term load forecasting: A comprehensive review and simulation study with CNN-LSTM hybrids approach. IEEE Access. 2024 Aug 8. DOI: 10.1109/ACCESS.2024.3440631

[16] Sajid M, Malik KR, Almogren A, Malik TS, Khan AH, Tanveer J, Rehman AU. Enhancing intrusion detection: a hybrid machine and deep learning approach. Journal of Cloud Computing. 2024 Jul 17;13(1):123. https://doi.org/10.1186/s13677-024-00685-x

[17] Afraji DM, Lloret J, Peñalver L. An integrated hybrid deep learning framework for intrusion detection in IoT and IIoT networks using CNN-LSTM-GRU architecture. Computation. 2025 Sep 14;13(9):222. https://doi.org/10.3390/computation13090222

[18] Wikle CK, Zammit-Mangion A. Statistical deep learning for spatial and spatiotemporal data. Annual Review of Statistics and Its Application. 2023 Mar 9;10(1):247-70. https://doi.org/10.1146/annurev-statistics-033021-112628

[19] Liu X, Shan J, Liu C, Zhang S, Zhang D, Hao Z, Huang S. An Operating Condition Diagnosis Method for Electric Submersible Screw Pumps Based on CNN-ResNet-RF. Processes. 2025 Jun 27;13(7):2043. https://doi.org/10.3390/pr13072043

[20] Bansal K, Singhrova A. Review on intrusion detection system for IoT/IIoT-brief study. Multimedia Tools and Applications. 2024 Mar;83(8):23083-108. https://doi.org/10.1007/s11042-023-16395-6

[21] Nascita A, Aceto G, Ciuonzo D, Montieri A, Persico V, Pescapé A. A survey on explainable artificial intelligence for internet traffic classification and prediction, and intrusion detection. IEEE Communications Surveys & Tutorials. 2024 Nov 22. DOI: 10.1109/COMST.2024.3504955

[22] Rekik S, Mehmood S. Hybrid GNN–LSTM defense with differential privacy and secure multi-party computation for edge-optimized neuromorphic autonomous systems. Scientific Reports. 2025 Dec 16;15(1):43939. https://doi.org/10.1038/s41598-025-27691-6

[23] Lilhore UK, Manoharan P, Simaiya S, Alroobaea R, Alsafyani M, Baqasah AM, Dalal S, Sharma A, Raahemifar K. Hidm: Hybrid intrusion detection model for industry 4.0 networks using an optimized cnn-lstm with transfer learning. Sensors. 2023 Sep 13;23(18):7856. https://doi.org/10.3390/s23187856

[24] Altunay HC, Albayrak Z. A hybrid CNN+ LSTM-based intrusion detection system for industrial IoT networks. Engineering Science and Technology, an International Journal. 2023 Feb 1; 38:101322. https://doi.org/10.1016/j.jestch.2022.101322

[25] Gupta H, Jadhav A, Bisht AS. Comparative Analysis of Machine and Deep Learning Models for Intrusion Detection in Fog-Enabled IoT Networks. International Journal of Networked and Distributed

Computing. 2026 Jun;14(1):1. https://doi.org/10.1007/s44227-025-00079-8

[26] Verma N, Kumar N, Almuzaini KK, Sinha A, Abid Hussain S. A real-time intelligent intrusion detection framework for robotic system cybersecurity. Peer-to-Peer Networking and Applications. 2026 Feb;19(1):30. https://doi.org/10.1007/s12083-025-02175-6

[27] Ogunseyi TB, Thiyagarajan G, He H, Bist V, Du Z. Performance Analysis of Explainable Deep Learning-Based Intrusion Detection Systems for IoT Networks: A Systematic Review. Sensors. 2026 Jan 6;26(2):363. https://doi.org/10.3390/s26020363

[28] Kalyani S, Vydeki D. A resource-efficient ensemble machine learning framework for detecting rank attacks in RPL-based IoT networks. Journal of Economy and Technology. 2026 Jan 1; 4:171-85. https://doi.org/10.1016/j.ject.2025.06.003

[29] Zahid M, Bharati TS. Leveraging Machine Learning and Deep Learning in IoT Security: A Review. Security and Privacy. 2026 Jan;9(1): e70144. https://doi.org/10.1002/spy2.70144

[30] Alasad Q, Ahmed M, Alahmed S, Khattab OT, Abdulwahhab SA, Yuan JS. A Comprehensive Review: The Evolving Cat-and-Mouse Game in Network Intrusion Detection Systems Leveraging Machine Learning. Journal of Cybersecurity and Privacy. 2026 Jan 4;6(1):13. https://doi.org/10.3390/jcp6010013