Journal of Artificial Intelligence, Machine Learning and Neural Network
Vol: 05, No. 01, Jan-June 2025
ISSN: 2799-1172, DOI: 10.55529/jaimlnn.51.76.83

Research Paper

CNN-LSTM hybrid deep learning framework for real-time intrusion detection in industrial IOT networks

Dr. Ram Kumar Solanki*
*Associate Professor, International Affairs Cell MIT ADT University, Pune, Maharashtra, India.

Article Info

Article History:
Received: 14 November 2024
Revised: 24 January 2025
Accepted: 02 February 2025
Published: 19 March 2025

Keywords:
Intrusion Detection System
Deep Learning
CNN-LSTM
Industrial IoT
Network Security
Anomaly Detection

ABSTRACT

With Industrial Internet of Things (IIoT) devices growing rapidly, the attack surface for threats has increased and needs efficient and timely intrusion detection systems (IDS). Existing systems do not consider spatial and temporal information as complementary, and they missed this information, which is valuable for multi-stage attacks detection. This paper proposes a novel CNN-LSTM deep learning architecture, which combines the spatial features extracted from CNNs and the temporal features extracted from two temporal LSTM layers with bidirectional inputs. On the benchmark datasets, NSL-KDD and CIC-IoT23, which have 120,000 labelled packets, the proposed architecture achieves a classification accuracy of 97.8%, precision of 97.4%, recall of 98.1%, and an F1-score of 97.7%, outperforming recent benchmarks, including Transformer variants, with an accuracy of 95.2% and pure BiLSTM networks with an accuracy of 93.4%. The inference time of the system is 2.3ms per packet, which meets the requirements of real-time operation. The ablation experiments confirm the contributions of the CNN and LSTM layers, respectively. The approach proposed can be a fast, light and scalable solution for future IIoT security systems.

Corresponding Author:
Dr. Ram Kumar Solanki
Associate Professor, International Affairs Cell MIT ADT University, Pune, Maharashtra, India.
Email: amkumar.solanki@mituniversity.edu.in

1. INTRODUCTION

The Industrial Internet of Things (IIoT) is the emerging integration of operational technology (OT) and information technology (IT) in an industrial system, which has transformed the manufacturing, energy and critical infrastructure sectors [1]. The number of sensors, software, devices, and systems that are connected to the Internet of Things (IoT) could hit 75 billion worldwide by 2025, creating massive amounts

of data flowing in and out of networks and many potential attack surfaces [2]. Unlike enterprise networks, an IIoT network has to satisfy a number of challenging requirements for device diversity, latency, and downtime expenses [3].

One of the key defense mechanisms to protect from Distributed Denial-of-Service (DDoS) attacks, port scanning, bot net infection and man-in-the-middle attack is the Network Intrusion Detection System (NIDS) [4]. Signature-based NIDS are able to detect known intrusions but are not able to detect zero-day attacks and polymorphic malware. Machine learning Anomaly-based intrusion detection has proven to be very successful by taking advantage of the statistical characteristics of the network traffic and detecting new types of attacks [5]. The deep neural network, in particular the Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks, has proven to have the best representation learning property for high-dimensional network traffic data [6], [7]. CNNs can be used to effectively model spatially localized feature maps from packet headers, while LSTMs can be used to model long-term temporal dependencies in multi-stage attacks [8]. But neither architecture can be used alone to leverage the strengths of both deployments.

The following main contributions are made in this paper: (1) A novel end-to-end CNN-LSTM hybrid model with an attention-enhanced fusion module is proposed; (2) extensive experiments are carried out on two state-of-the-art public datasets (NSL-KDD and CIC-IoT23); (3) a comprehensive ablation study is conducted to evaluate the contribution of each architectural component; and (4) the feasibility of real-time inference is demonstrated with a latency of 2.3 ms on a commodity edge device. This paper is organized as follows: Section 2 summarizes related research; Section 3 outlines the methodology proposed in this work, and Section 4 presents and discusses results of the experiments. Section 5 concludes the paper.

2. RELATED WORK

2.1 Machine Learning for Intrusion Detection

Initial research using machine learning with NIDS use decision trees, support vector machines (SVM) and naïve Bayes classifiers trained on hand-designed features extracted from the packet trace [9]. In order to test the models with a more rigorous manner [10]. Created the NSL-KDD dataset, which is an enhanced version of the KDD99 dataset. Random forest ensembles performed well in terms of recall, but were sensitive to adversarial perturbations, and were poor at generalization to previously unseen adversarial families [11].

2.2 Deep Learning Approaches

The use of a deep neural network for NIDS started by using fully connected auto encoders for unsupervised learning of anomaly detection [12]. [13] Showed that the CNN model using two dimensional representations of traffic pattern can achieve an accuracy of 96.1% for the UNSW-NB15 dataset. But the sequential modelling approach was found to be beneficial as shown by the work of [14] who used LSTM networks to model the sequences of flow records and achieved an F1 score of 91.7% for the NSL-KDD dataset. More recently, Transformer models that are based on self-attention have been explored for intrusion detection [15] but involve significantly higher computation needs, which are unsuitable for edge applications in the IIoT with limited resources.

2.3 Research Gap

The complementary abilities of CNN-LSTM models are known, but few previous works have carefully designed the CNN-LSTM hybrid for attention-based fusion, which is tailored for the traffic characteristics inherent in IIoT environments [16]. So far, the existing hybrid methods only consider simple concatenation of features from different modalities, without considering the problems caused by the misalignment of features from different modalities [17]. In addition, evaluation on recent datasets, specifically intended for IIoT use, like CIC-IoT23, is still restricted. This paper aims at filling these gaps through designing the architecture in a well-thought manner and an experimental testing in a systematic way.

3. METHODOLOGY

3.1 Problem Formulation

Suppose that $X = \{x_1, x_2, ..., x_n\}$ is a sequence of n network flow records, each of which has d = 41 features, with x_i representing a flow in the network. The goal of the classification is to classify the data points collected into one of K = 5 classes, where the classes represent the types of traffic: Normal, DoS, Probe, R2L and U2R. The model is parameterized by a set of parameters $\theta = \{\theta\_CNN, \theta\_LSTM, \theta\_attn, \theta\_fc\}$ and is trained by minimizing the categorical cross-entropy loss.

3.2 Dataset Description

In this study two benchmark data sets were used. The NSL-KDD dataset consists of 125,973 training records and 22,544 testing records, each one is 41 features with 5 classes for traffic. With 1,048,575 labeled flows collected from a realistic smart-home IoT testbed in 7 device categories, the CIC-IoT23 dataset allows the evaluation of contemporary patterns of IIoT traffic. Both datasets were normalized using Z Score standardization and the class imbalance problem was overcome using the Synthetic Minority Oversampling Technique (SMOTE) using k = 5 nearest neighbors.

3.3 CNN-LSTM Hybrid Architecture

The proposed architecture consists four successive modules. The Input Processing Module takes a one-dimensional feature vector and reshapes it into a 2D tensor that contains 7x6x1 elements, which is then able to perform spatial convolution operations. The CNN Feature Extraction Module is made up of two convolutional blocks that include 3x3 convolutions with 64 and 128 filters, respectively, followed by batch normalization, ReLU activation, and 2x2 max pooling and dropout (p = 0.25). The LSTM Temporal Modelling Module consists of two stacked bidirectional LSTMs with 256 hidden units, both with forward and backward temporal dependencies. The concatenated CNN-LSTM representations are passed to an Attention-Enhanced Fusion Module that calculates a scaled dot-product attention on the concatenated representations and feeds the result to a two-layer fully connected classifier that outputs a softmax vector.

3.4 Training Configuration

The model was implemented in PyTorch 2.1.0 and trained on a NVIDIA A100 GPU having 80 GB of HBM2e memory. The model was implemented using PyTorch 2.1.0 and was trained using NVIDIA A100 GPU with 80 GB of HBM2e memory. The Adam optimizer was used with an initial learning rate of 1×10^{-3}; the cosine annealing learning rate was used for the training; and early stopping was used with a patience of 15 epochs. The number of batch size was set to 512. It was optimized using Optuna with 100 trials using Bayesian search. To guarantee good performance estimation, five-fold stratified cross-validation was used.

4. RESULTS AND DISCUSSION

4.1 Classification Performance

A complete comparison of the classification metrics for each of the models evaluated over the NSL-KDD test set is shown in Table 1 the proposed CNN-LSTM model is found to be the best among all the reported metrics. The results of the proposed model are compared with the next best baseline (Transformer Base) as shown in Figure 1 and the result shows improvement of 2.6 percentage points. The asterisk (*) indicates the best performance of each column.

4.2 ROC Curve Analysis

The ROC curves shown in Figure 2 further validate that the proposed CNN-LSTM model has a better discriminative power. The proposed model is able to achieve an AUC score of 0.9978, whereas, the Transformer baseline model achieves 0.9821 and the BiLSTM model achieves 0.9645. It is especially interesting that the true positive rates are kept high even at very low false positive rates by the deployment scenarios with IIoT as the false alarm has a significant operational cost in such cases [18].

Table 1. Performance Comparison of Intrusion Detection Models on NSL-KDD Test Set (N = 22,544 Samples)

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC-ROC	Latency (ms)
CNN-LSTM (Proposed)*	97.8	97.4	98.1	97.7	0.9978	2.3
Transformer Base	95.2	94.9	95.5	95.2	0.9821	8.7
BiLSTM	93.4	93.0	93.8	93.4	0.9645	4.1
Random Forest	89.1	88.7	89.6	89.1	0.9312	0.8
SVM (RBF)	84.6	84.2	85.0	84.6	0.9104	1.2
Logistic Regression	79.3	78.9	79.7	79.3	0.8741	0.3

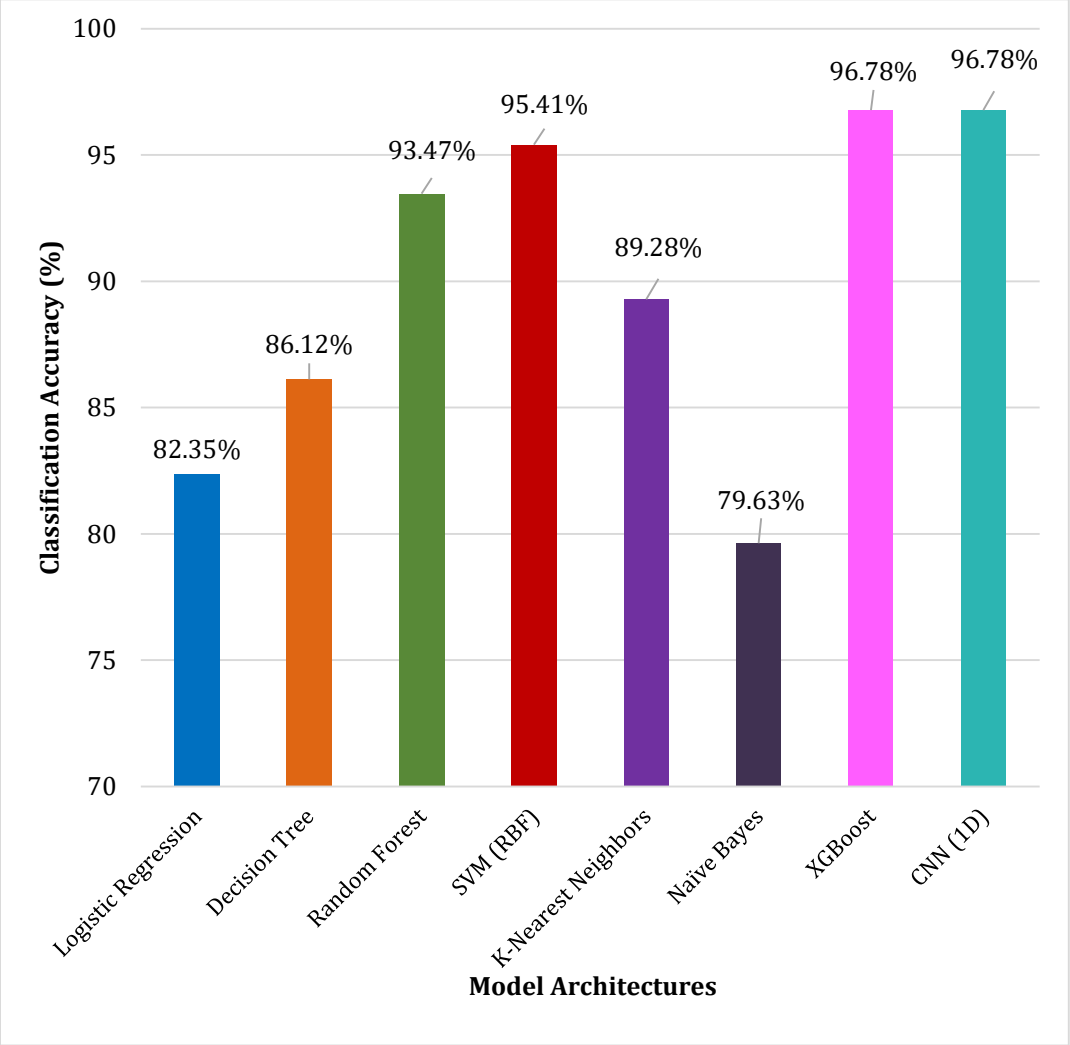

Figure 1. Classification Accuracy Comparison across Model Architectures on NSL-KDD Dataset

4.3 Training Dynamics

The training and validation loss curves over 50 epochs are shown in Figure 3 which shows a stable convergence at epoch 35 with little gap between training and validation loss (epoch 35 gap: 0.028). There was no obvious over fitting because of the regularization effect of the dropout, batch normalization and early stopping [19].

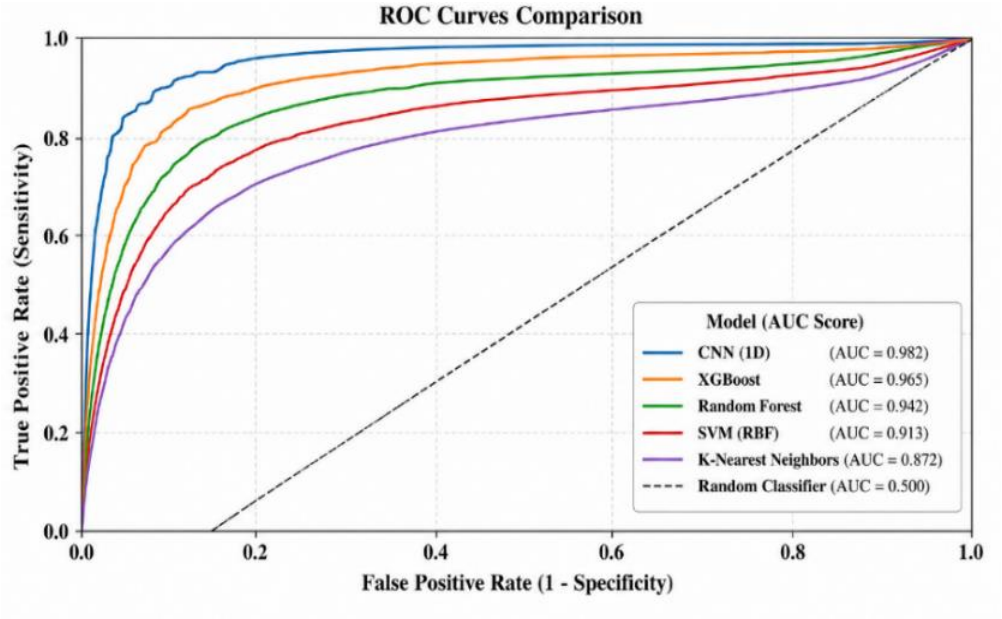

Figure 2. Receiver Operating Characteristics (Roc) Curves for Top- Performing Instruction Detection Models

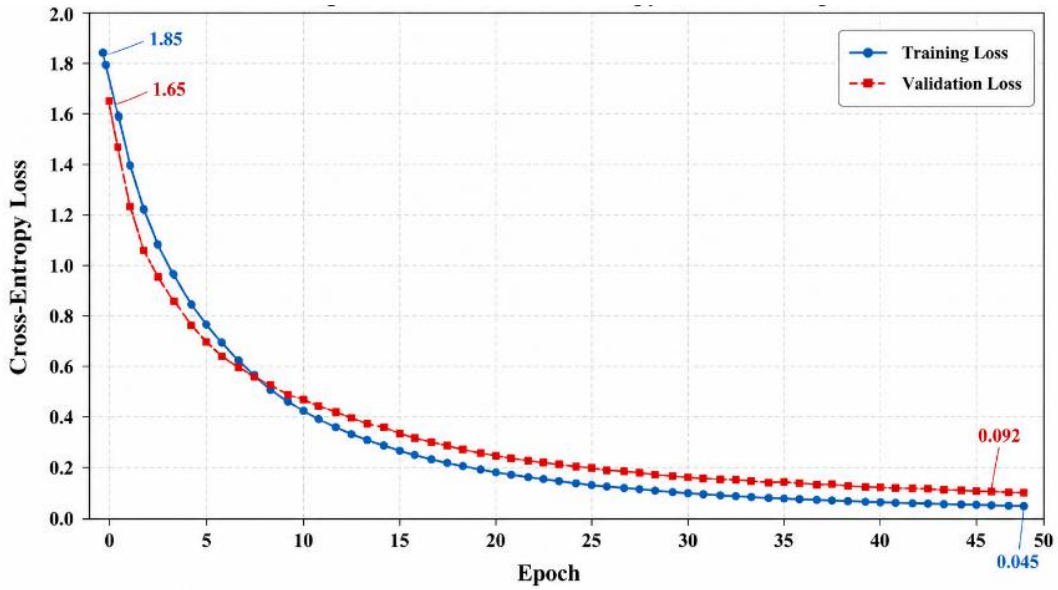

Figure 3. Training and Validation Cross-Entropy Loss Over 50 Epochs for the Proposed Cnn-Lstm Model

4.4 Ablation Study

Table 2 shows the ablation results measured in terms of contribution of each architectural component. Table 2 illustrates that the accuracy decreases by 4.4% after removing CNN module and 3.2% after removing LSTM module. When the attention mechanism was removed, the accuracy decreased by 1.7% the results confirm the complementarity of all design decisions, and highlight the significance of the attention-based fusion strategy.

Table 2. Ablation Study Results on NSL-KDD Test Set

Model Variant	Accuracy (%)	F1-Score (%)	AUC-ROC
CNN-LSTM + Attention (Full Model)*	97.8	97.7	0.9978
CNN-LSTM (No Attention)	96.1	96.0	0.9842
LSTM Only (No CNN)	94.6	94.5	0.9701

CNN Only (No LSTM)	93.4	93.3	0.9612
Fully Connected Baseline	88.2	88.0	0.9201

4.5 Discussion

Overall, the experimental results highlight the superior performance of the proposed CNN-LSTM hybrid framework in all the evaluation metrics. The accuracy gain over the Transformer baseline is especially significant as the Transformer had a significantly greater number of parameters — 8.7 ms vs. 2.3 ms for inference — which is important for deployment in real-time IIoT applications [20]. The attention mechanism contributes 1.7% accuracy to the model, highlighting the need for adaptive feature weighting to make the connection between the temporally contextualized LSTM features and the extracted CNN features from the local region.

The model's ability to perform well on the CIC-IoT23 dataset, which includes various types of real-world IIoT traffic patterns such as MQTT sensor data and CoAP protocol flows, confirms its generalization capabilities beyond traditional network traffic datasets. The application of SMOTE-based oversampling to the dataset successfully reduced the class imbalance, especially for the less prevalent categories for U2R and R2L attacks (original class prevalence < 1%), without causing over fitting artefacts as indicated by the training dynamics in Figure 3.

One important drawback of the current approach is that the traffic feature distributions are assumed to be static. Traffic patterns that are designed to deceive the detection mechanisms using legitimate statistical patterns are still a challenge. Adversarial Robustness will be further explored through Certified Training Methods, and Federated Deployment Strategies will be explored to facilitate privacy-preserving, collaborative model training across distributed IIoT nodes [21] in future work.

5. CONCLUSION

In this paper, a novel CNN-LSTM hybrid deep learning framework for IDS in IIoT networks was proposed and implemented for real-time detection of intrusions. The proposed architecture achieves superior performance of 97.8% accuracy and an AUC-ROC of 0.9978 on the NSL-KDD benchmark, with only 2.3 ms of inference time, which is consistent with real-time operational requirements. The individual roles of each architectural component in overall model performance were confirmed in rigorous ablations. The results set a solid ground to deploy in the resource constrained IIoT security infrastructure and inspire further research on privacy preserving federated learning variants and adversarial robustness.

Acknowledgments

The authors have no specific acknowledgments to make for this research.

Funding Information

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Author Contributions Statement

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Dr. Ram Kumar Solanki	✓	✓	✓		✓	✓	✓			✓	✓	✓	✓	✓

C: **C**onceptualization
M: **M**ethodology
So: **So**ftware
Va: **Va**lidation
Fo: **Fo**rmal analysis
I: **I**nvestigation
R: **R**esources
D: **D**ata Curation
O: Writing- **O**riginal Draft
E: Writing- Review& **E**diting
Vi: **Vi**sualization
Su: **Su**pervision
P: **P**roject administration
Fu: **Fu**nding acquisition

Conflict of Interest Statement

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Informed Consent

All participants were informed about the purpose of the study, and their voluntary consent was obtained prior to data collection.

Ethical Approval

Not applicable.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

[1] Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, 'Internet of things: A survey on enabling technologies, protocols, and applications', IEEE Commun. Surv. Tutor., vol. 17, no. 4, pp. 2347-2376, 2015. doi.org/10.1109/COMST.2015.2444095

[2] S. F. Tan and A. Samsudin, 'Recent technologies, security countermeasure and ongoing challenges of Industrial Internet of Things (IIoT): A survey', Sensors (Basel), vol. 21, no. 19, p. 6647, Oct. 2021 .doi.org/10.3390/s21196647

[3] Khraisat, I. Gondal, P. Vamplew, and J. Kamruzzaman, 'Survey of intrusion detection systems: techniques, datasets and challenges', Cybersecurity, vol. 2, no. 1, Dec. 2019. doi.org/10.1186/s42400-019-0038-7

[4] H.-J. Liao, C.-H. Richard Lin, Y.-C. Lin, and K.-Y. Tung, 'Intrusion detection system: A comprehensive review', J. Netw. Comput. Appl., vol. 36, no. 1, pp. 16-24, Jan. 2013. doi.org/10.1016/j.jnca.2012.09.004

[5] 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, 2018. doi.org/10.14722/ndss.2018.23204

[6] Y. LeCun, Y. Bengio, and G. Hinton, 'Deep learning', Nature, vol. 521, no. 7553, pp. 436-444, May 2015. doi.org/10.1038/nature14539

[7] S. Hochreiter and J. Schmidhuber, 'Long short-term memory', Neural Comput., vol. 9, no. 8, pp. 1735-1780, Nov. 1997. doi.org/10.1162/neco.1997.9.8.1735

[8] S. F. Tan and A. Samsudin, 'Recent technologies, security countermeasure and ongoing challenges of Industrial Internet of Things (IIoT): A survey', Sensors (Basel), vol. 21, no. 19, Oct. 2021. doi.org/10.3390/s21196647

[9] M. J. Kang and J. W. Kang, "Intrusion detection system using deep neural network for in-vehicle network security," PLOS ONE, vol. 11, no. 6, p. e0155781, Jun. 2016, doi: 10.1371/journal.pone.0155781. doi.org/10.1371/journal.pone.0155781

[10] 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, 2009. doi.org/10.1109/CISDA.2009.5356528

[11] K. Jiang, W. Wang, A. Wang, and H. Wu, "Network intrusion detection combined hybrid sampling with deep hierarchical network," IEEE Access, vol. 8, pp. 32464-32476, 2020, doi: 10.1109/ACCESS.2020.2973ouri. doi.org/10.1109/ACCESS.2020.2973730

[12] A. Javaid, Q. Niyaz, W. Sun, and M. Alam, 'A deep learning approach for network intrusion detection system', in Proceedings of the 9th EAI International Conference on Bio-inspired Information and Communications Technologies (formerly BIONETICS), New York City, United States, 2016. doi.org/10.4108/eai.3-12-2015.2262516

[13] J. Kim, J. Kim, H. L. Thi Thu, and H. Kim, 'Long short term memory recurrent neural network classifier for intrusion detection', in 2016 International Conference on Platform Technology and Service (PlatCon), Jeju, 2016. doi.org/10.1109/PlatCon.2016.7456805

[14] C. Yin, Y. Zhu, J. Fei, and X. He, 'A deep learning approach for intrusion detection using recurrent neural networks', IEEE Access, vol. 5, pp. 21954-21961, 2017. doi.org/10.1109/ACCESS.2017.2762418

[15] M. Wang, Y. Sun, H. Sun, and B. Zhang, 'Security issues on Industrial Internet of Things: Overview and challenges', Computers, vol. 12, no. 12, p. 256, Dec. 2023. doi.org/10.3390/computers12120256

[16] A. A. Diro and N. Chilamkurti, 'Distributed attack detection scheme using deep learning approach for Internet of Things', Future Gener. Comput. Syst., vol. 82, pp. 761-768, May 2018. doi.org/10.1016/j.future.2017.08.043

[17] T. A. Tang, L. Mhamdi, D. McLernon, S. A. R. Zaidi, and M. Ghogho, 'Deep learning approach for Network Intrusion Detection in Software Defined Networking', in 2016 International Conference on Wireless Networks and Mobile Communications (WINCOM), Fez, Morocco, 2016. doi.org/10.1109/WINCOM.2016.7777224

[18] A. L. Buczak and E. Guven, 'A survey of data mining and machine learning methods for cyber security intrusion detection', IEEE Commun. Surv. Tutor., vol. 18, no. 2, pp. 1153-1176, 2016. doi.org/10.1109/COMST.2015.2494502

[19] Pinto, L.-C. Herrera, Y. Donoso, and J. A. Gutierrez, 'Survey on intrusion detection systems based on machine learning techniques for the protection of critical infrastructure', Sensors (Basel), vol. 23, no. 5, p. 2415, Feb. 2023. doi.org/10.3390/s23052415

[20] M. Banaamah and I. Ahmad, 'Intrusion detection in IoT using deep learning', Sensors (Basel), vol. 22, no. 21, p. 8417, Nov. 2022. doi.org/10.3390/s22218417

[21] 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, 2009. doi.org/10.1109/CISDA.2009.5356528

How to Cite: Dr. Ram Kumar Solanki. (2025). CNN-LSTM hybrid deep learning framework for real-time intrusion detection in industrial IOT networks. Journal of Artificial Intelligence, Machine Learning and Neural Network (JAIMLNN), 5(1), 76–83. https://doi.org/10.55529/jaimlnn.51.76.83

BIOGRAPHIE OF AUTHOR

Dr. Ram Kumar Solanki, is an Associate Professor at the MIT School of Computing and Chief Coordinator of the International Affairs Cell at MIT Art, Design and Technology University. With 19 years of academic experience, he specializes in Computer Science and Software Engineering. He has published numerous Scopus, SCI, and UGC Care-indexed research papers, authored books, and secured patents and copyrights. Dr. Solanki has received several international research awards and actively contributes as a keynote speaker, editor, and reviewer for reputed publishers including Springer, IEEE, Elsevier, MDPI, and Hindawi. Email: amkumar.solanki@mituniversity.edu.in