UTM UNIVERSITI TEKNOLOGI MALAYSIA
INTERNATIONAL JOURNAL OF INNOVATIVE COMPUTING
ISSN 2180-4370
Journal Homepage : https://ijic.utm.my/

Anomaly Network-based Intrusion Detection Model Based on Hybrid CNN-LSTM and Attention Mechanism

Wang Tingyue[1*] & Maheyzah Md Siraj[2]
Faculty of Computing,
Universiti Tecknologi Malaysia,
81310, UTM Johor Bahru, Johor, Malaysia
Email: wangtingyue@graduate.utm.my[1], maheyzah@utm.my[2]

Submitted: 21/2/2025. Revised edition: 6/4/2025. Accepted: 11/4/2025. Published online: 27/5/2025
DOI: https://doi.org/10.11113/ijic.v15n1.524

***Abstract*—With the growing frequency of network attacks, traditional anomaly-based intrusion detection models often fail to identify advanced attack patterns and suffer from high false positive rates. This paper proposes a hybrid deep learning model integrating Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), and an Attention Mechanism to enhance detection accuracy and robustness. Leveraging CNNs for spatial feature extraction, LSTMs for temporal pattern recognition, and Attention Mechanisms for prioritizing critical data, the model effectively identifies diverse intrusion types. Using the NF-UNSW-NB15-v2 dataset, this research incorporates advanced preprocessing techniques such as Recursive Feature Elimination with Cross-Validation (RFECV) and Synthetic Minority Oversampling Technique (SMOTE). Experimental results demonstrate improved performance across key metrics, offering a robust framework for real-time intrusion detection in complex network environments.**

***Keywords*—Anomaly-Based Intrusion Detection Model (ANIDM) Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), Attention Mechanism, NF-UNSW-NB15-v2**

I. INTRODUCTION

The rapid growth of network applications has significantly increased the volume and complexity of network traffic, creating an ideal environment for cyberattacks [4]. These attacks range from Denial of Service (DoS) to sophisticated Advanced Persistent Threats (APTs), posing serious challenges to network security. Intrusion Detection Models (IDMs) have emerged as critical tools for identifying and mitigating such threats by monitoring network activity for abnormalities. However, traditional IDMs, especially those relying on signature-based techniques, are often unable to detect novel and evolving threats due to their dependency on predefined attack signatures. This limitation leads to high false positive rates and reduced detection accuracy in dynamic network environments.

The advent of deep learning has introduced powerful tools for addressing these challenges. Deep learning models, particularly Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks, offer advanced capabilities in feature extraction and temporal pattern recognition [5], [6]. CNNs excel at identifying spatial features in structured data, while LSTMs are designed to capture dependencies over time, making them well-suited for analyzing sequential network traffic data. Despite their individual strengths, combining these models has shown potential to achieve even greater detection accuracy by leveraging their complementary capabilities.

Attention Mechanisms further enhance this hybrid approach by dynamically prioritizing the most relevant features in the data [9]. This reduces noise and improves the interpretability and robustness of the model. The integration of CNN, LSTM, and Attention Mechanisms offers a comprehensive framework for detecting anomalies in network traffic, addressing limitations in existing models.

This paper proposes a novel anomaly-based IDM leveraging this hybrid approach. Using the NF-UNSW-NB15-v2 dataset, a highly diverse and realistic network traffic dataset, the model incorporates advanced preprocessing techniques such as RFECV for feature selection and SMOTE for data balancing. The proposed model is evaluated using various performance metrics, including accuracy, precision, recall, and

F1-score, demonstrating its effectiveness in handling diverse attack types.

II. RELATED WORKS

Intrusion Detection Models (IDMs) are proactive cybersecurity systems that monitor network activity in real-time, issuing alerts or taking actions when potential intrusions are detected. Based on data sources, IDMs can be classified into Host-based IDM (HIDM), which monitors individual devices, and Network-based IDM (NIDM), which analyzes network traffic. Alternatively, detection strategies divide IDMs into anomaly-based, signature-based, and misuse-based approaches, each with distinct advantages and limitations [1], [2]. While anomaly-based systems excel in detecting novel attack patterns, they suffer from high false positive rates, especially in dynamic environments [3]. Signature and misuse detection methods, though accurate for known attacks, struggle with scalability and evolving threats [4]. Despite advancements, current Anomaly Network-Based Intrusion Detection Models (ANIDMs) face challenges such as adaptability to dynamic environments, handling encrypted traffic, ensuring high accuracy with imbalanced data, and scalability for big data processing [5], [6]. This research addresses these limitations through a hybrid CNN-LSTM-Attention model, emphasizing improved feature extraction, reduced false positives, and scalable big data processing for robust anomaly detection.

Research on Anomaly Network-Based Intrusion Detection Models (ANIDMs) has advanced significantly, addressing issues like high false positive rates, low detection rates, and class imbalances. Support Vector Machines (SVMs) have been widely used, achieving notable accuracies in various studies, but often face challenges in scalability and adaptability to evolving threats [7], [8], [9]. Deep learning approaches, such as CNNs and LSTMs, have shown promising results in enhancing detection accuracy, with models integrating skip connections and attention mechanisms achieving high precision rates [10], [11]. Despite these advancements, challenges remain in handling diverse attack types, managing large-scale network traffic, and ensuring real-world applicability, particularly in IoT environments [12]. This research builds on prior work by combining CNN, LSTM, and Attention Mechanisms to address these limitations, enabling more comprehensive spatial and temporal feature extraction and improving overall detection robustness.

Deep Learning (DL), a subfield of Machine Learning (ML), excels in studying the intrinsic patterns of data and has achieved remarkable success in areas such as text, image, and sound recognition, surpassing traditional techniques [13], [14]. By leveraging multi-layered neural networks, DL enables machines to mimic human activities like hearing, thinking, and pattern recognition, driving advancements in Artificial Intelligence (AI) [15], [16]. Despite its advantages, including strong learning capabilities, wide adaptability, and compatibility with frameworks like TensorFlow and PyTorch [17], [18], DL faces significant challenges. These include high computational and hardware costs, reliance on extensive datasets, and the complexity of designing new models, which demand substantial resources and expertise [19][20]. Addressing these limitations is crucial for expanding the practical applications of DL, especially in resource-constrained environments. The following next is review on Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM) and Attention Mechanism.

A. Review on CNN

Convolutional Neural Network (CNN), inspired by biological vision systems, are key deep learning models excelling in computer vision tasks like object detection and image generation [21]. Unlike traditional neural networks, CNN achieve translation invariance by employing convolution operations, enabling them to capture local features in images regardless of their position [22]. Through convolutional layers, CNN utilize specific filters or kernels to extract diverse image features, making them fundamental to modern computer vision and deep learning research [23]. Fig. 1 shows the basic structure of CNN.

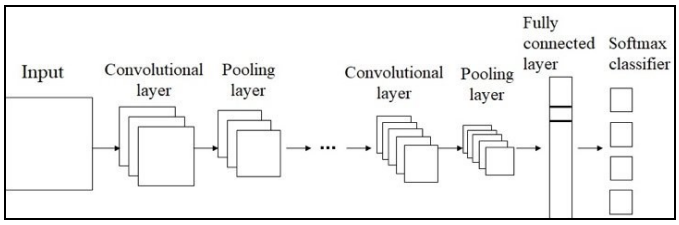

Fig. 1. Basic Structure of CNN [24]

B. Review on LSTM

Long Short-Term Memory (LSTM) network are a specialized form of Recurrent Neural Network (RNN) designed to capture long-term dependencies in sequential data. By incorporating mechanisms such as forget gates, input gates, and output gates, LSTM dynamically retain important information while discarding irrelevant data, addressing the limitations of traditional RNN in handling long-term dependencies [25]. Forget gates determine which components of prior states to discard, while input gates control the integration of new information into the current cell state, and output gates decide the final output based on the updated memory [26]. These mechanisms allow LSTM to effectively manage the flow of information, solving issues like gradient vanishing and explosion in long-sequence training [27]. Fig. 2 shows the internal sructure of LSTM.

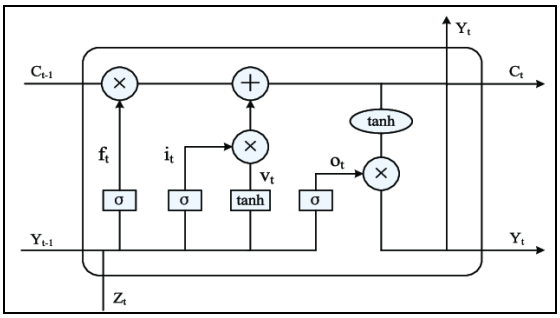

Fig. 2. Internal Structure of LSTM [26]

C. Review on Attention

The Attention Mechanism enhances neural networks by focusing on the most relevant parts of input sequences, mimicking human attention. By dynamically assigning different weights to sequence elements, it prioritizes critical information while reducing noise. This mechanism, utilizing components like Query, Key, and Value, is widely applied in tasks such as natural language processing and sequence modelling, improving the performance and efficiency of deep learning models [28], [29]. Fig. 3 shows Attention Mechanism.

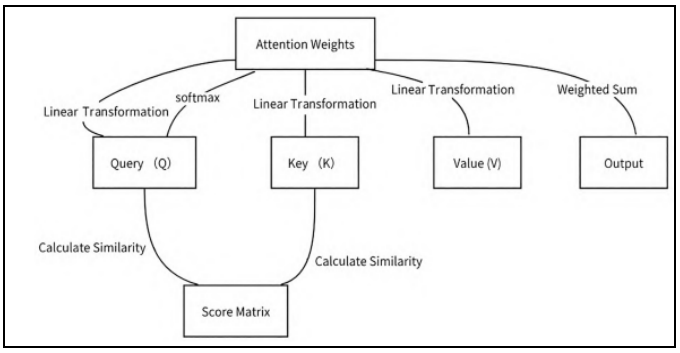

Fig. 3. Attention Mechanism [29]

The hybridization of CNN, LSTM, and Attention Mechanism provides a robust solution to the multifaceted challenges in intrusion detection. CNN extracts spatial features from structured data, while LSTM captures temporal dependencies, making it suitable for analyzing sequential network traffic. The Attention Mechanism further enhances the model by dynamically prioritizing critical features, improving interpretability and accuracy. Together, these components create a comprehensive system capable of handling diverse and complex intrusion patterns, reducing false positives, and improving classification accuracy in real-world scenarios.

The NF-UNSW-NB15-v2 dataset is selected for its modern attack diversity, extensive size, and realistic traffic representation. It ensures a comprehensive evaluation of the hybrid model's performance across varied attack types. This research focuses on addressing challenges like encrypted traffic, data imbalance, and dynamic network conditions by integrating advanced deep learning techniques, aiming to enhance detection accuracy, efficiency, and adaptability to emerging threats in cybersecurity.

Literature Review establishes the foundation for a robust hybrid intrusion detection model by reviewing key methodologies, challenges, and datasets, and justifying the integration of CNN, LSTM, and Attention Mechanism.

III. THE PROPOSED METHODOLOGY

The methodology adopted in this research to develop a robust hybrid intrusion detection model, which is structured into three distinct phases: data pre-processing, model design, training and evaluation. Each phase addresses specific challenges in intrusion detection, ensuring a systematic approach to achieving high detection accuracy and robustness. Fig. 4 shows the framework of the proposed model.

The NF-UNSW-NB15-v2 dataset was used in this study due to its comprehensive and realistic representation of modern network traffic. Developed by the Australian Centre for Cyber Security (ACCS), this dataset contains over 2.5 million labeled records, categorized into ten distinct classes including various attack types and benign traffic. Each record is composed of more than 45 flow-based features such as protocol types, byte counts, packet counts, and timing characteristics. These features make the dataset particularly suitable for spatial and temporal analysis, enabling the effective application of CNN and LSTM layers. Moreover, its imbalanced class distribution poses a realistic challenge, making it ideal for evaluating the robustness and generalization capability of intrusion detection models.

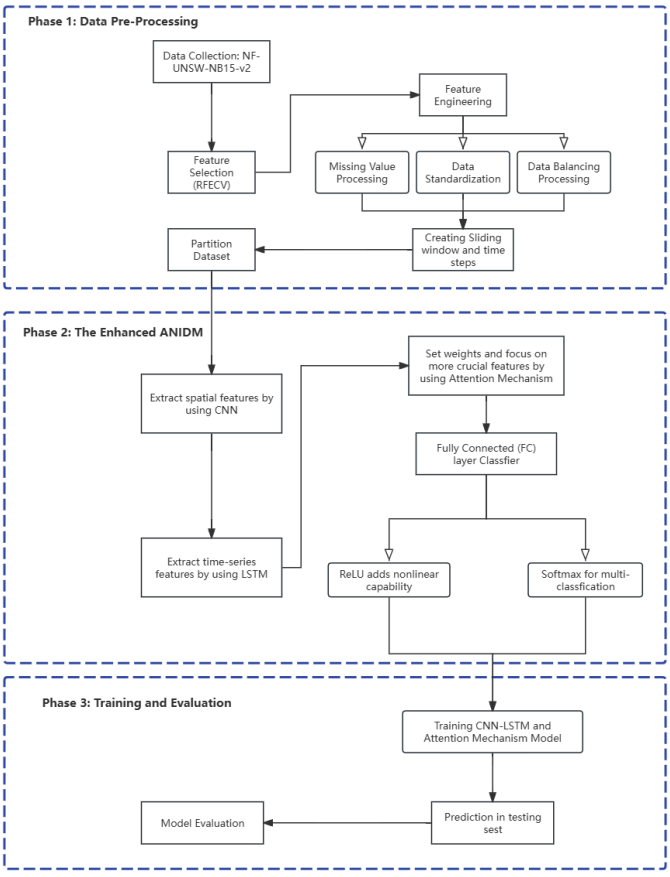

Fig. 4. The Proposed Research Framework

Data pre-processing focuses on preparing the NF-UNSW-NB15-v2 dataset for model training. Key steps include feature selection using Recursive Feature Elimination with Cross-Validation (RFECV) to identify the most relevant attributes, addressing data imbalance with the Synthetic Minority Oversampling Technique (SMOTE), and creating sliding windows to capture temporal dependencies in network traffic. These steps ensure that the dataset is both representative and optimized for hybrid model training.

The hybrid model integrates CNN for spatial feature extraction, LSTM for capturing temporal dependencies, and the Attention Mechanism for prioritizing critical features. CNN layers detect localized patterns in network traffic, LSTM layers model the sequential nature of traffic flows, and the Attention Mechanism dynamically emphasizes relevant features, improving classification accuracy and reducing noise. This phase establishes a cohesive architecture to handle the complexities of intrusion detection.

The model undergoes rigorous training using the Adam optimizer with a dynamically adjusted learning rate over 100 epochs. Performance is evaluated using metrics such as accuracy, precision, recall, F1-score, and ROC/PR curves, with additional visualizations like confusion matrices to analyze class-wise performance. These comprehensive evaluations ensure the model's reliability and ability to generalize to unseen data.

The phased approach ensures that the hybrid model effectively integrates spatial, temporal, and prioritized feature extraction, demonstrating its capability to enhance detection accuracy and robustness in dynamic network environments.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

A. Data Pre-processing

The design and setup of the research environment focus on tools like MySQL for managing the NF-UNSW-NB15-v2 dataset, Anaconda for virtual environments, and PyCharm for development. These tools ensure an efficient and seamless pipeline for handling large datasets and building the hybrid CNN-LSTM-Attention model.

The data collection process highlights the use of the NF-UNSW-NB15-v2 dataset, selected for its diversity in attack types and realistic network traffic scenarios. The dataset is stored in MySQL, enabling effective preprocessing and querying to prepare it for model training.

The preprocessing phase addresses critical steps like feature selection using *RFECV*, normalization with *MinMaxScaler*, and balancing classes with *SMOTE*. Temporal dependencies are captured using sliding windows, and the data is partitioned for fair model evaluation. This comprehensive pipeline ensures high-quality inputs for robust intrusion detection model performance. Table I shows the selected features after data pre-processing.

TABLE I. SELECTED FEATURES AND DESCRIPTIONS

Features	Descriptions
L4_SRC_PORT	ipv4 source port number
L4_DST_PORT	ipv4 destination port number
PROTOCOL	IP protocol identifier byte
L7_PROTO	Layer 7 protocol (numeric)
IN_BYTES	Incoming number of bytes
IN_PKTS	Incoming number of packets
DNS_QUERY_ID	DNS query transaction Id
OUT_BYTES	Outgoing number of bytes
OUT_PKTS	Outgoing number of packets
TCP_FLAGS	Cumulative of all TCP flags
CLIENT_TCP_FLAGS	Cumulative of all client TCP flags
SERVER_TCP_FLAGS	Cumulative of all server TCP flags
FLOW_DURATION_MILLISECONDS	Flow duration in milliseconds
DNS_QUERY_TYPE	DNS query type (e.g. 1=A, 2=NS)
DURATION_IN	Client to Server stream duration (msec)
MIN_TTL	Min flow TTL
MAX_TTL	Max flow TTL
LONGEST_FLOW_PKT	Longest packet (bytes) of the flow
SHORTEST_FLOW_PKT	Shortest packet (bytes) of the flow
MIN_IP_PKT_LEN	Len of the smallest flow IP packet observed
DNS_TTL_ANSWER	TTL of the first A record (if any)
MAX_IP_PKT_LEN	Len of the largest flow IP packet observed
SRC_TO_DST_SECOND_BYTES	Src to dst Bytes/sec
DST_TO_SRC_SECOND_BYTES	Dst to src Bytes/sec
RETRANSMITTED_IN_BYTES	Number of retransmitted TCP flow bytes (src->dst)
RETRANSMITTED_IN_PKTS	Number of retransmitted TCP flow packets (src->dst)
RETRANSMITTED_OUT_BYTES	Number of retransmitted TCP flow bytes (dst->src)
RETRANSMITTED_OUT_PKTS	Number of retransmitted TCP flow packets (dst->src)
SRC_TO_DST_AVG_THROUGHPUT	Src to dst average thpt (bps)
DST_TO_SRC_AVG_THROUGHPUT	Dst to src average thpt (bps)
NUM_PKTS_UP_TO_128_BYTES	Packets whose IP size <= 128
NUM_PKTS_128_TO_256_BYTES	Packets whose IP size > 128 and <= 256
NUM_PKTS_256_TO_512_BYTES	Packets whose IP size > 256 and <= 512
NUM_PKTS_512_TO_1024_BYTES	Packets whose IP size > 512 and <= 1024
NUM_PKTS_1024_TO_1514_BYTES	Packets whose IP size >1024 and <= 1514
TCP_WIN_MAX_IN	Max TCP Window (src->dst)
TCP_WIN_MAX_OUT	Max TCP Window (dst->src)
ICMP_TYPE	ICMP Type * 256 + ICMP code
ICMP_IPV4_TYPE	ICMP Type

B. Model Design

The hybrid CNN-LSTM-Attention model is meticulously designed to tackle the multifaceted challenges of intrusion detection by leveraging the complementary strengths of its three main components: CNN, LSTM, and Attention Mechanism. CNN plays a critical role in extracting spatial features from structured network traffic data, identifying localized patterns that may signify anomalies. LSTM complements this by analyzing temporal dependencies, effectively modeling sequential behaviors and capturing long-term dependencies in time-series data. The Attention Mechanism further refines the model's focus by dynamically prioritizing the most relevant features, enabling it to emphasize critical data points while reducing noise and improving interpretability.

The design process, as depicted in Fig. 5, provides a comprehensive representation of the model's operational flow, starting from data preprocessing to feature extraction, temporal analysis, and final classification. During data preprocessing, steps such as feature selection, normalization, and class

balancing are employed to optimize input quality. The CNN layers then extract and condense spatial features, passing them to LSTM layers that model the temporal evolution of the data. The Attention layer assigns varying weights to different features, ensuring the model allocates its computational resources effectively to improve accuracy.

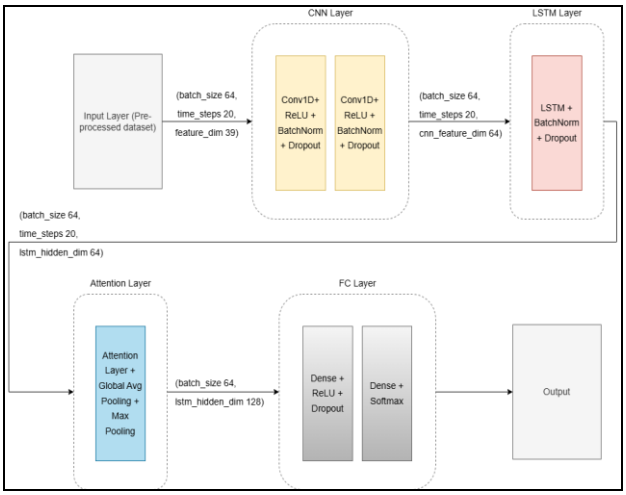

Fig. 5. The Proposed CNN-LSTM-Attention Model

Table II provides a detailed summary of the CNN-LSTM-Attention model architecture. It outlines the layers used, their respective output shapes, and the number of trainable parameters. The model incorporates convolutional layers for spatial feature extraction, LSTM for temporal feature capture, and an Attention Mechanism for focusing on critical input features. Additionally, the training configuration is specified at the end of the table.

TABLE II. PROPOSED CNN-LSTM-ATTENTION ARCHITECTURE

Layer (type)	Output Shape	Param
input_3 (Input Layer)	(None, 20, 39)	0
conv1d_4 (Conv1D)	(None, 20, 32)	3776
batch_normalization_6	(None, 20, 32)	128
dropout_8 (Dropout)	(None, 20, 32)	0
conv1d_5 (Conv1D)	(None, 20, 64)	6208
batch_normalization_7	(None, 20, 64)	256
dropout_9 (Dropout)	(None, 20, 64)	0
lstm_2 (LSTM)	(None, 20, 64)	24832
batch_normalization_8	(None, 20, 64)	256
dropout_10 (Dropout)	(None, 20, 64)	0
dense_10, dense_11, dense_12 (Dense)	(None, 20, 64)	12480
layer_normalization_2	(None, 20, 64)	128
global_average_pooling1d_2 (GlobalAveragePooling1D)	(None, 64)	0
global_max_pooling1d_2 (GlobalMaxPooling1D)	(None, 64)	0
tf. concat_2 (Concatenate)	(None, 128)	0
dense_13 (Dense)	(None, 32)	4128
dropout_11 (Dropout)	(None, 32)	0
dense_14 (Dense)	(None, 10)	330
Total Params		52,522
Trainable Params		52,202
Non-trainable Params		320
Validation Split		0.2
Verbose		1

Table III describes the main parameter configurations of the CNN-LSTM-Attention model. The model takes time series data as input, extracts local features through CNN, then learns the dependencies between sequences through LSTM, and combines Attention mechanism to further focus on the feature contribution of key time steps. The Global Pooling layer (Global Average and Max Pooling) combines the global features and finally completes the classification through the full connection layer. This table details the design details of the model input dimension, the configuration of the convolutional and LSTM layers, the attention weight dimension, the output dimension of the fully connected layer, and the activation function.

TABLE III. CNN-LSTM-ATTENTION MODEL PARAMETERS

S. NO	Parameters	Value
1	Enter data dimension	(20, 29)
2	Number of convolutional layers	2
3	Convolution layer output dimension	(20, 64)
4	Convolution kernel size	3
5	Pooling layer size	GlobalAveragePooling1D() GlobalMaxPooling1D()
6	Number of LSTM layers	1
7	Number of LSTM units	64
8	Attention weight dimension	7
9	Output dimension of the FC Layer	32
10	Activation function	ReLU, Softmax

Table IV lists the range of hyperparameters of CNN-LSTM-Attention model in the training process. Adam was selected as the optimization algorithm, and the balance between convergence speed and stability was achieved by adjusting the learning rate range (1e-6 to 0.0005). Batch size increased from 64 to 128 to speed up training; The number of training iterations is set to 50 to 100; L2 regularization coefficients (0.001 to 0.01) limit model complexity to prevent overfitting; The Dropout ratio (0.2 to 0.3) is used to randomly drop some neurons to improve the model's generalization ability. These hyperparameter Settings ensure model stability and performance optimization.

TABLE IV. MODEL HYPERPARAMETERS

S. NO	Hyperparameters	Minimum	Maximum
1	Learning rate	1e-6	0.0005
2	Batch size	64	128
3	Number of training iterations	50	100
4	Regularization coefficient	0.001	0.01
5	Optimization algorithm	Adam	N/A
6	Dropout rate	0.2	0.3

C. Performance Evaluation

The CNN-LSTM-Attention model demonstrates strong generalization and robustness, with accuracy steadily improving and stabilizing at 98.4% on the test set and 98.2%

on the training set. Simultaneously, the loss values for both sets decrease consistently throughout the training process, stabilizing around 0.05 as shown in Fig. 6. These results, illustrated in the combined accuracy and loss curve, highlight the model's balanced fit and effective handling of complex data, with regularization techniques mitigating overfitting and ensuring reliable performance.

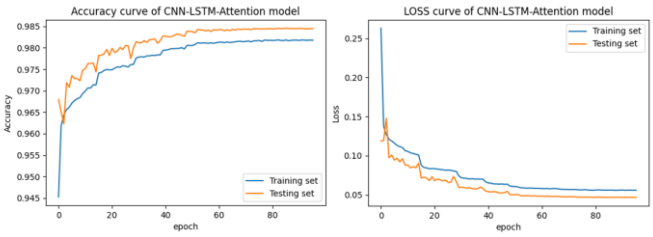

Fig. 6. Accuracy and Loss Curve of Proposed Model

The confusion matrix analysis, Fig. 7, shows that most categories are classified accurately, with Class 2 ‘Benign’ and Class 9 ‘Worms’ achieving nearly 100% accuracy. However, some misclassifications occur, such as Class 3 ‘DoS’ being confused with Class 4 ‘Exploits’ and Class 6 ‘Generic’ with Class 7 ‘Reconnaissance’, likely due to overlapping feature distributions. Larger sample categories, like Class 0 ‘Analysis’ and Class 1 ‘Backdoor’ perform better, while smaller categories exhibit higher error rates, reflecting the impact of class imbalance on model performance.

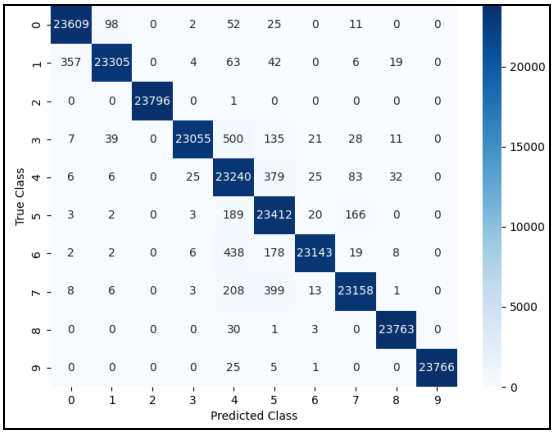

Fig. 7. Confusion Matrix

As shown in Fig 8, the combined ROC and PR curve analysis demonstrates the CNN-LSTM-Attention model's exceptional classification performance across all categories. With both AUC and average precision (AP) values at 1.00, the model shows a remarkable ability to balance high precision and recall rates under various thresholds. These results, consistent with the confusion matrix and F1 scores, confirm that the model achieves near-zero error rates for most categories, effectively handling both high-frequency and low-frequency classes. The model's robust architecture, integrating CNN for feature extraction, LSTM for temporal modeling, and Attention for feature weighting, is key to its superior performance and generalization in multi-class anomaly detection tasks.

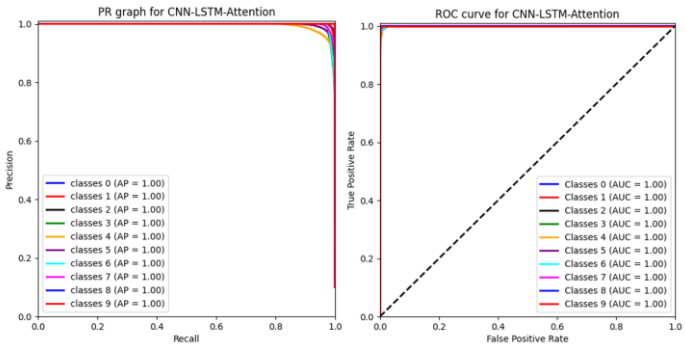

Fig. 8. ROC and PR Graph of Proposed Model

The CNN-LSTM-Attention model demonstrates excellent performance in multi-class anomaly detection as shown in Table V, achieving 100% accuracy, precision, recall, and F1 score for the Benign class, and near-perfect F1 scores for classes like Analysis (98.81%), Backdoor (98.64%), and Shellcode (99.78%). While slightly lower scores are observed for DoS (98.33%), Exploits (95.75%), Fuzzers (96.80%), and Reconnaissance (97.99%) due to feature complexity and overlaps, the model's overall metrics remain high, with an average accuracy, precision, recall, and F1 score of 98.44%, 98.48%, 98.44%, and 98.45%, respectively, showcasing its robustness and reliability across diverse categories.

TABLE V. PROPOSED MODEL MULTI-CLASSIFICATION RESULTS

	Accuracy	Precision	Recall	F1-Score
Analysis	99.21%	98.40%	99.21%	98.81%
Backdoor	97.94%	99.35%	97.94%	98.64%
Benign	100.00%	100.00%	100.00%	100.00%
DoS	96.89%	99.81%	96.89%	98.33%
Exploits	97.66%	93.91%	97.66%	95.75%
Fuzzers	98.39%	95.26%	98.39%	96.80%
Generic	97.26%	99.64%	97.26%	98.43%
Reconnaissance	97.32%	98.67%	97.32%	97.99%
Shellcode	99.86%	99.70%	99.86%	99.78%
Worms	99.87%	100.00%	99.87%	99.93%
Overall	**98.44%**	**98.48%**	**98.44%**	**98.45%**

D. Comparative Analysis

Table VI shows that the proposed CNN-LSTM-Attention model significantly outperforms existing methods across all key metrics. Compared to the Transformer-CNN model [30], it achieves notable improvements, with accuracy increasing by 3.08%, precision by 4.33%, recall by 3.08%, and F1 score rising from 0.8860 to 0.9845. Against MyCNN [31] and mCNN [32] models, the proposed model shows dramatic gains, particularly in precision, which improves from 0.4530 and 0.3570 to 0.9848, underscoring the critical role of the Attention

Mechanism in optimizing feature selection and classification accuracy. Even when compared to the closely performing RandomForest model [33], the proposed model achieves higher recall (0.9844) and F1 score (0.9845), demonstrating its robust and reliable performance in handling diverse anomaly detection tasks. These results highlight the effectiveness of integrating CNN, LSTM, and Attention Mechanism for advanced intrusion detection.

TABLE VI. COMPARISON WITH OTHER MODELS

Model	Accuracy	Precision	Recall	F1-Score
Transformer-CNN [30]	0.9537	0.8515	0.9537	0.8860
MyCNN [31]	0.9937	0.4530	0.7710	0.4580
mCNN [32]	0.9858	0.3570	0.6310	0.3710
RandomForest [33]	0.9609	0.94	0.95	0.94
Proposed Model	**0.9844**	**0.9848**	**0.9844**	**0.9845**

V. CONCLUSION

The study presents a comprehensive hybrid model for anomaly-based intrusion detection, integrating CNN, LSTM, and Attention Mechanisms. By addressing challenges like class imbalance, complex attack detection, and high false positive rates, experimental results in Section IV demonstrate that the proposed CNN-LSTM-Attention model outperforms all compared methods, including traditional models like Random Forest and deep learning models such as MyCNN, mCNN, and Transformer-CNN. It achieved the highest accuracy (98.44%), precision (98.48%), recall (98.44%), and F1-score (98.45%), confirming its robustness and effectiveness in handling multi-class anomaly detection tasks. Future directions include exploring the model's scalability for encrypted traffic and its deployment in cloud-based environments, paving the way for advanced network security solutions.

ACKNOWLEDGMENT

I would like to express my sincere gratitude to my supervisor, Dr. Maheyzah, for her valuable guidance and support throughout the research process. Her expertise and insights were invaluable in shaping my research and helping me to overcome challenges.

CONFLICTS OF INTEREST

The author(s) declare(s) that there is no conflict of interest regarding the publication of this paper.

REFERENCES

[1] Leevy, J. L., & Khoshgoftaar, T. M. (2020). A Survey and Analysis of Intrusion Detection Models based on cse-cic-ids2018 Big Data. *Journal of Big Data*, *7*, 1–19.

[2] Martins, I., Resende, J. S., Sousa, P. R., Silva, S., Antunes, L., & Gama, J. (2022). Host-based IDS: A Review and Open Issues of an Anomaly Detection System in IoT. *Future Generation Computer Systems*, *133*, 95–113.

[3] Alsoufi, M. A., Razak, S., Siraj, M. M., Nafea, I., Ghaleb, F. A., Saeed, F., & Nasser, M. (2021). Anomaly-Based Intrusion Detection Systems in IoT using Deep Learning: A Systematic Literature Review. *Applied Sciences*, *11*(18), 8383.

[4] Masdari, M., & Khezri, H. (2020). A Survey and Taxonomy of the Fuzzy Signature-based Intrusion Detection Systems. *Applied Soft Computing*, *92*, 106301.

[5] Bhatti, U. A., Huang, M., Wu, D., Zhang, Y., Mehmood, A., & Han, H. (2019). Recommendation System using Feature Extraction and Pattern Recognition in Clinical Care Systems. *Enterprise Information Systems*, *13*(3), 329–351.

[6] Riyad, A. M., Ahmed, M. I., & Khan, R. R. (2019). An Adaptive Distributed Intrusion Detection System Architecture using Multi Agents. *International Journal of Electrical and Computer Engineering (IJECE)*, *9*(6), 4951–4960.

[7] Guezzaz, A., Benkirane, S., & Azrour, M. (2022). A Novel Anomaly Network Intrusion Detection System for Internet of Things Security. *IoT and Smart Devices for Sustainable Environment* (pp. 129–138). Cham: Springer International Publishing.

[8] Liu, Z., Thapa, N., Shaver, A., Roy, K., Yuan, X., & Khorsandroo, S. (2020). Anomaly Detection on IoT Network Intrusion using Machine Learning. *2020 International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems (icABCD)* (pp. 1–5). IEEE.

[9] Zhang, Y., Yang, Q., Lambotharan, S., Kyriakopoulos, K., Ghafir, I., & AsSadhan, B. (2019). Anomaly-based Network Intrusion Detection using SVM. *2019 11th International Conference on Wireless Communications and Signal Processing (WCSP)* (pp. 1–6). IEEE.

[10] Al-Turaiki, I., & Altwaijry, N. (2021). A Convolutional Neural Network for Improved Anomaly-based Network Intrusion Detection. *Big Data*, *9*(3), 233–252.

[11] Fu, Z. (2022). Computer Network Intrusion Anomaly Detection with Recurrent Neural Network. *Mobile Information Systems*, *2022*(1), 6576023.

[12] Saba, T., Rehman, A., Sadad, T., Kolivand, H., & Bahaj, S. A. (2022). Anomaly-based Intrusion Detection System for IoT Networks through Deep Learning Model. *Computers and Electrical Engineering*, *99*, 107810.

[13] Kelleher, J. D. (2019). *Deep Learning*. MIT Press.

[14] Buduma, N., Buduma, N., & Papa, J. (2022). *Fundamentals of Deep Learning*. O'Reilly Media, Inc.

[15] Christin, S., Hervet, É., & Lecomte, N. (2019). Applications for Deep Learning in Ecology. *Methods in Ecology and Evolution*, *10*(10), 1632–1644.

[16] Thompson, N. C., Greenewald, K., Lee, K., & Manso, G. F. (2020). The Computational Limits of Deep Learning. *arXiv preprint arXiv:2007.05558*.

[17] Wang, X., Han, Y., Leung, V. C., Niyato, D., Yan, X., & Chen, X. (2020). Convergence of Edge Computing and Deep Learning: A Comprehensive Survey. *IEEE Communications Surveys & Tutorials*, *22*(2), 869–904.

[18] Weng, C., Shah, N. H., & Hripcsak, G. (2020). Deep Phenotyping: Embracing Complexity and Temporality—Towards Scalability, Portability, and Interoperability. *Journal of Biomedical Informatics*, *105*, 103433.

[19] Chen, J., & Ran, X. (2019). Deep Learning with Edge Computing: A Review. *Proceedings of the IEEE*, *107*(8), 1655–1674.

[20] Qian, S., Pham, V. H., Lutellier, T., Hu, Z., Kim, J., Tan, L., ... & Shah, S. (2021). Are My Deep Learning Systems Fair?

An Empirical Study of Fixed-seed Training. *Advances in Neural Information Processing Systems*, *34*, 30211–30227.

[21] Kattenborn, T., Leitloff, J., Schiefer, F., & Hinz, S. (2021). Review on Convolutional Neural Network (CNN) in Vegetation Remote Sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, *173*, 24–49.

[22] Li, Z., Liu, F., Yang, W., Peng, S., & Zhou, J. (2021). A Survey of Convolutional Neural Network: Analysis, Applications, and Prospects. *IEEE Transactions on Neural Networks and Learning Systems*, *33*(12), 6999–7019.

[23] Yang, A., Wang, L. J., Ma, W. N., Tang, M., & Chen, J. (2022). Shape from Shading-based Study of Silica Fusion Characterization Problems. *Minerals*, *12*(10), 1286.

[24] You, Wei & Shen, Changqing & Wang, Dong & Chen, Liang & Jiang, Xingxing & Zhu, Zhongkui. (2019). An Intelligent Deep Feature Learning Method with Improved Activation Functions for Machine Fault Diagnosis. *IEEE Access.* PP. 1–1.

[25] Sherstinsky, A. (2020). Fundamentals of Recurrent Neural Network (RNN) and Long Short-term Memory (LSTM) Network. *Physica D: Nonlinear Phenomena*, *404*, 132306.

[26] Zeng, Y., Chen, J., Jin, N., Jin, X., & Du, Y. (2022). Air Quality Forecasting with Hybrid LSTM and Extended Stationary Wavelet Transform. *Building and Environment*, *213*, 108822.

[27] Song, X., Liu, Y., Xue, L., Wang, J., Zhang, J., Wang, J., ... & Cheng, Z. (2020). Time-series Well Performance Prediction based on Long Short-term Memory (LSTM) Neural Network Model. *Journal of Petroleum Science and Engineering*, *186*, 106682.

[28] Niu, Z., Zhong, G., & Yu, H. (2021). A Review on the Attention Mechanism of Deep Learning. *Neurocomputing*, *452*, 48–62.

[29] Cheng, Y., Yao, L., Xiang, G., Zhang, G., Tang, T., & Zhong, L. (2020). Text Sentiment Orientation Analysis based on Multi-channel CNN and Bidirectional GRU with Attention Mechanism. *IEEE Access*, *8*, 134964–134975.

[30] Kamal, H., & Mashaly, M. (2024). Advanced Hybrid Transformer-CNN Deep Learning Model for Effective Intrusion Detection Systems with Class Imbalance Mitigation Using Resampling Techniques. *Future Internet, 16*(12), 481.

[31] Sayed, N., Shoaib, M., Ahmed, W., Qasem, S., Albarrak, A., & Saeed, F. (2022). Augmenting IoT Intrusion Detection System Performance using Deep Neural Network. *Computers, Materials and Continua*, *74*(1), 1351–1374.

[32] Ajagbe, S. A., Awotunde, J. B., & Florez, H. (2023). Ensuring Intrusion Detection for IoT Services through an Improved CNN. *SN Computer Science*, *5*(1), 49.

[33] Ba, A., & Adda, M. (2024). Intrusion Detection in IIoT Using Machine Learning. *Procedia Computer Science*, 251, 265–272.