Journal of Theoretical and Applied Information Technology
31st July 2025. Vol.103. No.14

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

ATTENTION-ENHANCED LSTM MODEL FOR INTRUSION DETECTION IN IMBALANCED NETWORK TRAFFIC DATA

YAKUB REDDY.K , G.SHANKARLINGAM

[1]Research Scholar, Department of Computer Science and Engineering, Chaitanya Deemed to be university, Warangal, Telangana, India

[2]Professor, Department of Computer Science and Engineering, Chaitanya deemed to be University Warangal, Telangana, India.

yakubreddy1245@gmail.com , shankar@chaitanya.edu.in

ABSTRACT

Intrusion Detection Systems (IDS) face significant challenges in identifying minority attack classes within imbalanced network traffic, leading to compromised security in critical systems. To address this, we propose an Attention-Enhanced Long Short-Term Memory (AE-LSTM) model that integrates multi-head attention mechanisms with Long Short-Term Memory (LSTM) networks for robust intrusion detection. The model is trained on the NSL-KDD dataset using a comprehensive preprocessing pipeline that includes one-hot encoding, normalization, and SMOTE-based oversampling to mitigate class imbalance, particularly for rare attack types such as User-to-Root (U2R) and Remote-to-Local (R2L). Our architecture incorporates an LSTM layer with multi-head attention, residual connections, and dense layers with dropout regularization. Experimental results demonstrate a classification accuracy of 98.43% and a Top-5 accuracy of 100%**.** ROC-AUC scores reached **1.00** for most classes, and Precision-Recall analysis confirmed high sensitivity for minority attacks. Visualization via t-SNE revealed distinct inter-class separation. The proposed AE-LSTM model significantly enhances detection performance on imbalanced datasets, presenting a promising approach for next-generation intrusion detection systems (IDS).

Keywords: *Intrusion Detection, LSTM, Attention, SMOT, NSL-KDD*

1. INTRODUCTION

In today's digital world, the complexity and vulnerability of network infrastructures have significantly increased. Cybersecurity has become a paramount concern, with intrusion detection systems (IDS) playing a critical role in safeguarding digital assets against different activities. The other IDS approaches, often relying on authentication-based detection, struggle to identify novel and sophisticated attacks, necessitating the adoption of more advanced methodologies.

Machine learning (ML) has emerged as a powerful tool for handling statistical data. In networking, intrusion detection, and cyber attacks, as per [2] and [16], ML models can capture non-sequence patterns and classify data. However, dynamic and sequence data features are interdependent, and samples also depend on each other. Therefore, these ML models will not capture sequential patterns from the networking data. Although this model provides better accuracy, it cannot handle adversarial attacks. Instead of simple machine learning (ML) methods, some researchers, such as [16], have implemented ensemble methods like AdaBoost and Random Forests, which can be adapted to handle class imbalance by focusing on instances that are difficult to classify.

Deep learning (DL), a subset of machine learning (ML), addresses these limitations by automatically learning hierarchical feature representations from raw data. Among DL architectures, LSTM networks from [4], [13], [14], and [15] have shown promise in modeling sequential data, capturing temporal dependencies essential for analyzing network traffic patterns. LSTM networks are particularly effective in handling time-series data; they are perfectly suitable for intrusion detection tasks because this data consists of many dependency features related to time and attack. However, only a sequential model, such as LSTM, can capture temporal

sequential patterns from a large amount of data when handling it. And the CNN model will not capture sequential dependencies from the given samples.

A significant challenge while using ML and DL-based methods for IDS is the issue of class imbalance in network traffic datasets. Typically, datasets contain a disproportionate number of regular traffic instances compared to various types of attacks, including sporadic but critical ones such as U2R and R2L attacks. This imbalance in classes can bias the learning process, which will lead to reduced performance for minority classes.

When implementing a model for multiple-class classification, class imbalances can create models that are biased towards the majority class, leading to high overall accuracy but poor performance in detecting minority class instances. This is particularly problematic in intrusion detection, where the minority classes often represent the most critical threats and have a dynamic nature. And the attackers will attack every time with rare patterns, so there is no chance to delete minority classes. For instance, a model might achieve high accuracy by correctly identifying regular traffic but fail to detect rare but important attacks, rendering the IDS ineffective. Without handling the minority classes, training any model is not applicable.

To address class imbalance, various techniques have been proposed to mitigate this issue. One common approach is data-level methods, which involve resampling the dataset to balance the class distribution. Oversampling techniques, such as the Synthetic Over Sampling Technique, generate synthetic samples for minority classes, thereby balancing the dataset and giving equal priority to all classes, allowing for more effective learning from underrepresented classes. Another approach is algorithm-level methods, which modify the learning algorithm to reduce bias towards the majority class. This includes cost-sensitive learning, where misclassification costs are adjusted to penalize errors in minority classes more heavily. In addition to handling class imbalance, incorporating attention mechanisms into deep learning models has improved intrusion detection capabilities. The attention method in LSTM will take position-level inputs and sequences related to time. Integrating attention with LSTM networks enables a more nuanced analysis of networking and flow, which improves the complexity and accuracy of the model in detection.

2. RELATED WORK

Recent advancements in IDS, threat detection, and cybersecurity analytics have increasingly incorporated DL, hybrid architectures, and optimization techniques. Several studies have explored these integrations using benchmark and contemporary datasets to address evolving security challenges in both traditional and IoT-driven environments. Sharma et al. [1] introduced MA-Deep, a multi-attention-based deep convolutional recurrent neural network, which was trained on the CICIoMT 2024 dataset—a modern and dynamically evolving IoT dataset. Their model achieved a remarkable accuracy of 99.49%, benefiting from the synergy between spatial feature extraction through CNNs and temporal modeling via RNNs. However, the authors acknowledged that CNNs, while effective for spatial patterns, fall short in capturing positional and sequential dependencies inherent in network traffic. Kanthimathi et al. [2] developed multiple hybrid models that combine CNNs with Salp Swarm Optimization (SSO), followed by classifiers such as XGBoost, LSTM, and Random Forest. Using the CICDDoS2019 dataset, the proposed models achieved a maximum accuracy of 98.63%, highlighting the efficacy of optimization and ensemble techniques for DDoS detection.

Bhattacharya et al. [3] employed a CNN-RNN hybrid model using the NSL-KDD dataset. Despite achieving a modest accuracy of 81.38%, their findings indicated that while hybrid models offer promise, RNNs suffer from vanishing gradient issues, limiting their ability to capture long-term dependencies within sequential network traffic data. To address these limitations, Imrana et al. [4] proposed a CNN-GRU-FF model, which was tested on both the NSL-KDD and UNSW-NB15 datasets. Achieving an impressive 99.69% accuracy, the study demonstrated that gated recurrent units (GRUs), when combined with feedforward layers, effectively capture sequential and contextual patterns in intrusion data. Sukanya et al. [5] implemented a hybrid intrusion detection system (IDS) that fuses anomaly detection with conventional classification techniques using the UNSW-NB dataset. Although the model achieved an accuracy of 85.10%, it highlighted the importance of unsupervised pre-filtering in enhancing detection sensitivity. However, this method is limited in its ability to counter evolving adversarial attacks due to static clustering

mechanisms. Ganapaneni et al. [6] explored standalone and hybrid models—CNN, RNN, and CNN-RNN—on the NSL-KDD dataset. Consistent with Bhattacharya et al. [3], their models also reported an accuracy of 81.38%, underscoring the constraints of traditional datasets and the need for advanced feature engineering. Cui et al. [7] focused on intelligent feature extraction to improve IDS performance. Using the ToN_IoT and UNSW-NB15 datasets, their model achieved 99.55% accuracy, indicating that insufficient feature extraction can compromise robustness and suggesting the need for automated, intelligent pipelines.

Aljabri et al. [8] introduced MHAID-IWSOA, an IDS that utilizes multi-headed attention enhanced by an Improved Whale Swarm Optimization Algorithm. Tested on the Edge-IIoT dataset, it achieved 98.28% accuracy, showcasing the benefits of combining attention mechanisms with evolutionary optimization for edge computing environments. Logeswari et al. [9] developed a hybrid approach integrating Adaptive Neuro-Fuzzy Inference Systems (ANFIS), RNN, and Quantum-Inspired Particle Swarm Optimization (QIPSO) on the ToN-IoT and BOT-IoT datasets. Their model achieved 98.60% accuracy, validating the potential of combining fuzzy logic and deep learning (DL) to handle uncertainty in IoT data. Mao et al. [10] proposed MFEI-IDS, a Multi-Feature Extraction and Integration model using fully connected layers, evaluated on ISCX 2012 and CIC-IDS 2017 datasets. With 99.47% accuracy, their work demonstrated the strength of fusing diverse features from multiple input streams to improve generalizability. Oladele et al. [11] introduced ABA-IDS, tailored for Controller Area Network (CAN) traffic. This anomaly-based model achieved 99.30% accuracy, highlighting the growing importance of IDS in automotive networks and embedded communication systems. Vadisetty et al. [12] addressed the issue of adversarial machine learning in Intrusion Detection Systems (IDS). Their model, tested across multiple datasets (UNSW_NB15, CICIDS2017, NID, and TON_IoT), achieved an accuracy of 98.90%. Their findings highlighted the critical need for IDS robustness against adversarial samples in critical infrastructure. Wu et al. [13] implemented an Enhanced Residual-MBi-LSTM model on the HighD dataset. Achieving 98.01% accuracy, the combination of residual learning and bi-directional LSTM proved beneficial for anomaly detection in vehicular time-series data. Awan et al. [14] proposed SACNN, a spatial attention-based CNN designed for malware detection using the Malimg dataset. The model yielded 98.62% accuracy, demonstrating how attention mechanisms can enhance spatial localization in image-based malware classification. Liao et al. [15] developed a Multi-Channel Fusion model integrated with a Convolutional Block Attention Module (MCF-CBAM), which achieved the highest accuracy among all reviewed works, at 99.94%. This model, tested on N-BaIoT, KDDCUP99, and UNSW-NB15, underscored the value of cross-channel attention in enhancing learning granularity. Immadisetty et al. [16] revisited classical machine learning approaches, employing XGBoost and SVM for anomaly classification on the KDD dataset. Despite achieving a respectable 94.20% accuracy, the study reinforced the viability of traditional machine learning (ML) techniques for resource-constrained environments. Pham et al. [17] proposed the AAGCN (Attention-Augmented Graph Convolutional Network) for activity recognition, evaluating it on the CMDFALL, MICA-Action3D, and NTU-RGBD datasets. Reporting 88.2% accuracy, they highlighted the promise of graph-based models in capturing spatial-temporal relations in human behavior analysis.

Zhang et al. [18] introduced a novel traffic representation approach for intrusion detection using the NSL-KDD dataset, achieving an accuracy of 95.20%. The study emphasized learning representations directly from raw network traffic without relying on handcrafted features. Chithanuru et al. [19] implemented a hybrid RAE-GAMI-Net framework in Proof-of-Stake blockchain environments. With 96.20% accuracy, the study underscored the growing need for anomaly detection within decentralized systems. Alharbi et al. [20] presented a hybrid model combining YAMNet and CNNs for industrial fault detection (IFD), evaluated on the MIMII dataset. Achieving an accuracy of 83.55%, the study addressed the challenges of acoustic-based fault detection in noisy environments. Qiao et al. [21] employed a Bi-LSTM architecture on a custom cattle video dataset to detect behavioral anomalies, achieving an accuracy of 90.70%. This work extended IDS applications to smart agriculture and livestock monitoring. Chao et al. [22] developed MST3D, a multi-scale spatiotemporal 3D convolutional model for behavioral anomaly prevention, achieving an accuracy of 90.94%. Their work showcased the

utility of 3D Convolutional Neural Networks (CNNs) in surveillance-based anomaly detection. Jiang et al. [23] introduced memory-guided feature learning for visual anomaly detection, which was evaluated on the UCSD Ped2 and CUHK datasets. With 99.50% accuracy, the model demonstrated the effectiveness of attention-enhanced memory units in dense, context-rich environments. Malik et al. [24] compared CNN and SVM models for emotion-based anomaly detection using CASME II and SAMM datasets. The hybrid model achieved an accuracy of 89.40%, reflecting the value of combining deep learning and conventional classifiers for affective analysis. Abd El-Nabi et al. [25] employed the Fast Gradient Sign Method (FGSM) for adversarial training on Eye-Blink and CEW datasets, attaining 99.82% accuracy. However, the study noted that while high-performing, the model lacks robustness against evolving adversarial threats in vision-based IDS, particularly for human-computer interaction.

3. METHODOLOGY

We implemented a hybrid model that combines LSTM with Multi-Head Attention mechanisms to effectively capture sequential data, such as time series or natural language, for classification tasks. This integration will provide the strengths of both LSTM and attention mechanisms to capture temporal dependencies and focus on salient features within the input sequences.

The model begins with an input layer that accepts data shaped according to the input shape, typically representing sequences over time with multiple features. This input is fed into an LSTM layer model consisting of 128 units, and the return_sequences parameter is set to True to preserve the temporal structure for subsequent layers. The LSTM layer also has dropout and recurrent dropout rates of 0.2 to prevent overfitting by randomly deactivating a fraction of units during training. LSTM networks are adept at capturing long-term dependencies in sequential data due to their gated architecture, which regulates the flow of information and gradients through time.

After the LSTM layer, the trained data is passed to a Multi-Head Attention mechanism with four heads and a key dimension of 32. This layer allows the model to attend to different positions within the sequence simultaneously, enabling it to capture various contextual relationships. The attention process updates the weights with a sequence, considering the relevance of each time step to highlight essential features that contribute to effective prediction. The output of the attention method is combined with the original LSTM output through a residual connection, and the result is normalized using Layer Normalization with a small epsilon value of 1e-6 to maintain numerical stability.

After the sequential layers, a feed-forward layer is added, consisting of a dense layer with 128 units and an activation function. With this, all vectors are converted to the range [0 to +α]. A Dropout layer with a dropout rate of 0.3 follows this thick layer. It means that 30% of the neurons are dropped from the network after every iteration. This combination introduces non-linearity and further regularization, enhancing the model's ability to learn complex patterns perfectly with neurons. The final output is passed through a Global Average Pooling layer, which aggregates the information across the time dimension, resulting in a fixed-size vector that summarizes the sequence's salient features.

Finally, the model concludes with a dense output layer employing a softmax activation function, producing a probability distribution over the target classes. This configuration is suitable for multi-class classification problems, where the model predicts the likelihood of each class given the input sequence.

a. Data pre-processing and Normalization

The NSL-KDD dataset from Kaggle is used to train the proposed model for evaluating Intrusion Detection Systems (IDS). It consists of various features, including many classes. However, the dataset inherently suffers from class imbalance, where certain attack types, such as U2R and R2L, have a minimal number of samples compared to other classes, like DoS attacks and regular traffic. This class imbalance in data will create several challenges, such as the potential for overfitting, which may lead to bias towards both the majority and minority classes.

To address these challenges, a comprehensive data preprocessing strategy was employed. Initially, categorical features such as protocol type, service, and flag were transformed using one-hot encoding, converting them into numerical representations suitable for machine learning algorithms. This step ensures that the model can effectively interpret these categorical variables.

Subsequently, normalization techniques were applied to scale numerical features. Normalization is essential because the features in

the dataset vary widely in scale, ranging from one-digit numbers to five-digit numbers; for instance, the duration feature can range from 0 to over 58,000 seconds. Without normalization, features with larger scales will dominate the learning process, leading to a biased model. The minority classes are balanced by generating synthetic samples of the minority classes through interpolation between existing minority instances, as shown in Figure 1. After preprocessing and balancing, the dataset was prepared for training and testing. The training set consisted of 100,778 cases, while the testing set comprised 25,195 instances.

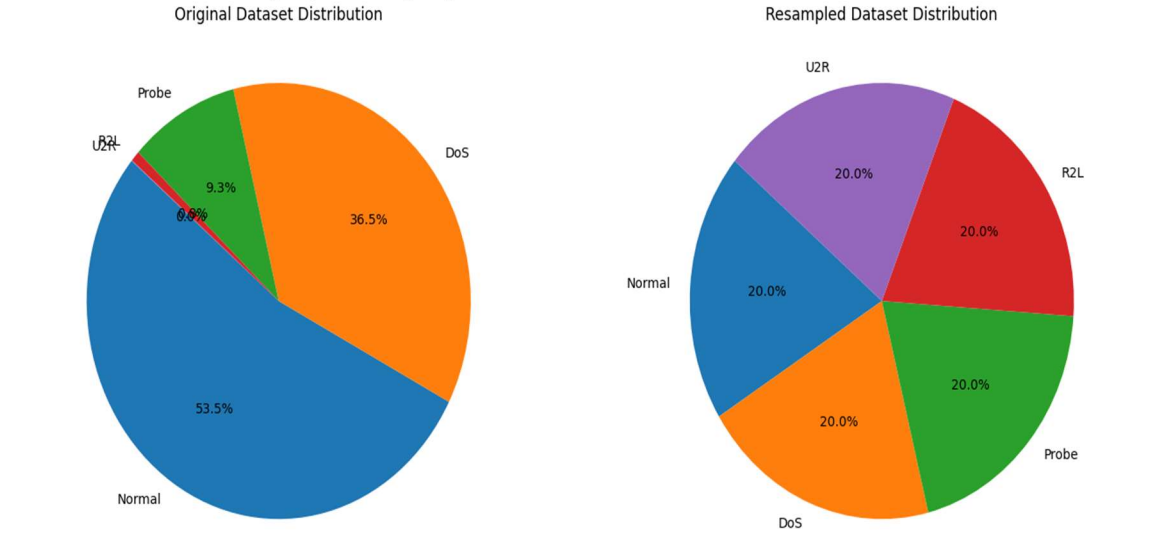

Figure 1: Original And Pre-Processed Data

4. RESULT ANALYSIS AND DISCUSSION

The training process for the proposed deep learning model was conducted over 10 epochs, utilizing a T5 GPU to leverage its high computational capabilities. The model demonstrated significant performance improvements throughout the training period, indicating effective learning and generalization.

In the first epoch with random weights, the model achieved an accuracy of 64.01% and 83.50% on the training and validation sets, respectively, with corresponding loss values of 0.9031 and 0.4634. This early stage reflects the model's adaptation to the data and the commencement of learning complex patterns.

However, from the second epoch, the training and validation accuracy increased to 78.09% and 94.01%, respectively, accompanied by a decrease in training loss to 0.5205 and validation loss to 0.1998. These enhancements suggest that the model is effectively capturing intricate features and reducing error rates. By the third epoch, with suitable weights, the model achieved training and validation accuracies of 80.98% and 96.90%, respectively. From this, it is observed that, with the random weights and the Adam optimizer, the weights are correctly updated, and the training and validation losses further decrease to 0.4286 and 0.0968, respectively. This indicates continued refinement in feature extraction and model optimization.

Throughout the subsequent epochs, the model's performance improves epoch by epoch, reaching training and validation accuracies of 96.72% and 98.34% by the tenth epoch, as illustrated in Figure 2. The training and validation losses also decreased correspondingly, indicating the model's robustness in learning complex features. In this, the accuracy and loss are correctly updated, preventing overfitting.

The training durations per epoch ranged from approximately 361 to 439 seconds, reflecting the computational demands of processing a large dataset with complex features. Despite the substantial time investment, the consistent improvement in accuracy and loss metrics underscores the model's efficacy and the adequacy of the computational resources employed.

Journal of Theoretical and Applied Information Technology
31st July 2025. Vol.103. No.14

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

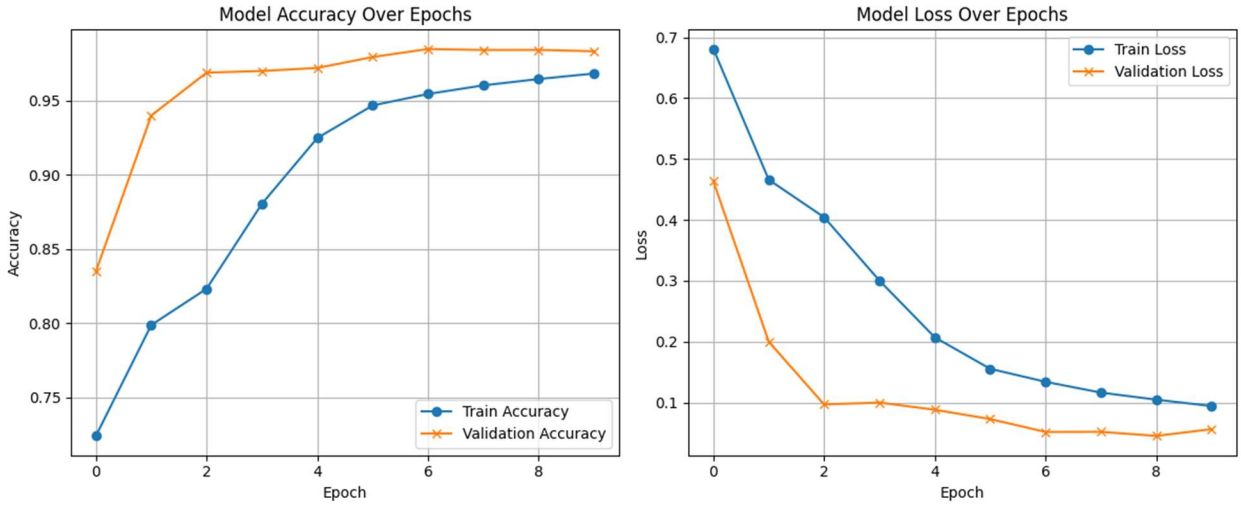

Figure 2: Learning Curves Of Proposed Models

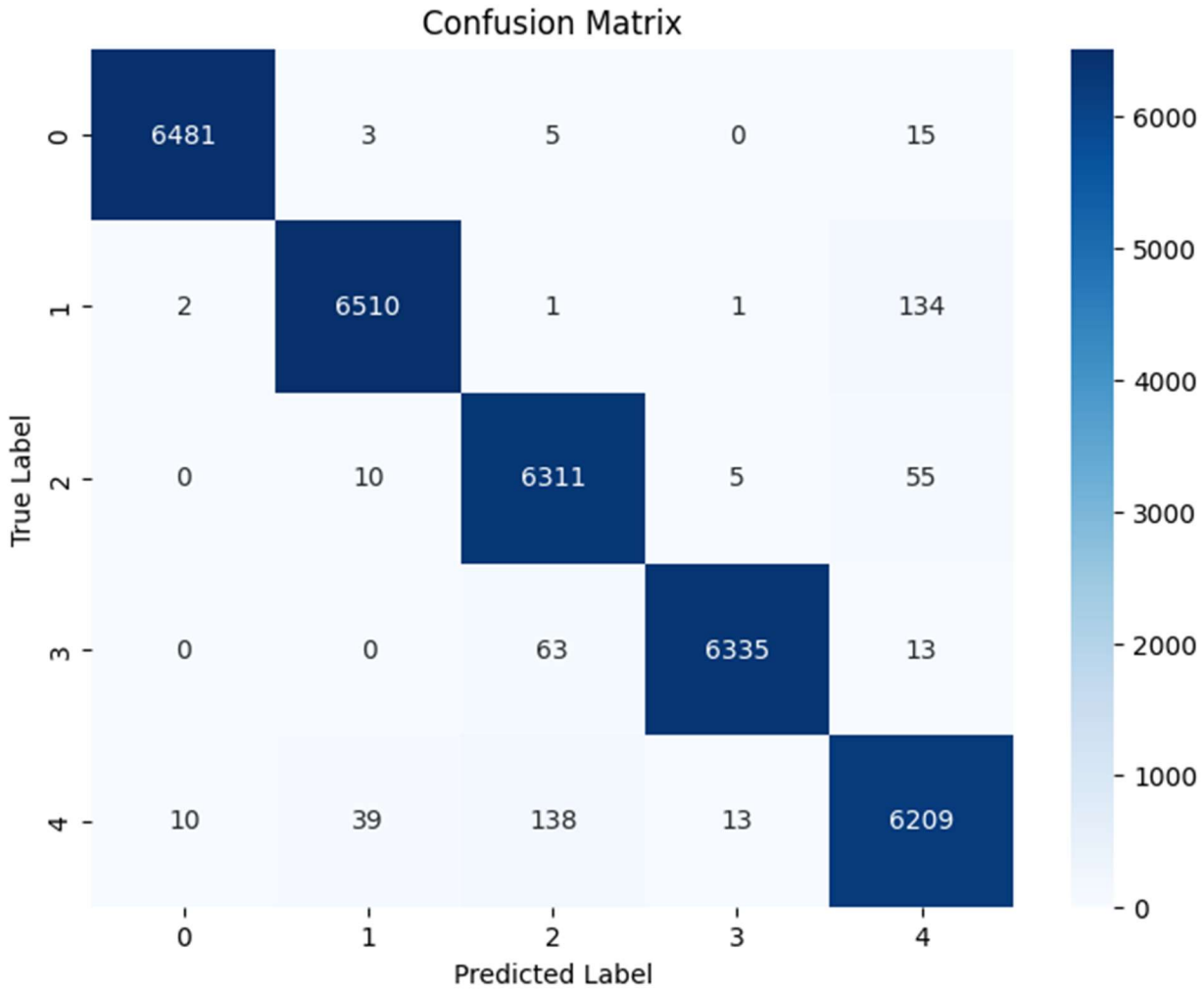

Figure 3: Confusion Matrix Of The Proposed Model

Journal of Theoretical and Applied Information Technology
31st July 2025. Vol.103. No.14

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

The proposed deep learning-based intrusion detection model demonstrates exceptional performance across all classes, effectively addressing both common and rare attack types. The evaluation metrics reveal that the model accurately identifies instances from each class, including the minority classes, indicating its robustness and reliability.

Class 0 (Normal): The model achieved near-perfect accuracy in classifying regular traffic, with a very high rate of correct identifications and a very low rate of misclassifications. This suggests that the model is highly effective in distinguishing between normal and abnormal network activities.

Class 1 (DoS): For Denial of Service (DoS) attacks, the model demonstrated a strong ability to identify attacks while minimizing false positives.

Class 2 (Probe): In detecting probing activities, the model maintained a high rate of correct classifications, ensuring that probing attempts are accurately identified.

Class 3 (R2L): The model demonstrated a strong capability in identifying Remote-to-Local attacks, accurately classifying instances with minimal errors.

Class 4 (U2R): Despite the rarity of User-to-Root attacks, the model effectively identified these instances, demonstrating its ability to detect even the least frequent attack types with high accuracy, as shown in Figure 3.

The model achieved an overall accuracy of 98.43%, indicating that it correctly classified a significant majority of the instances in the test dataset, as illustrated in Figure 4.

Top-N Accuracy: The model's top-N accuracy metrics further demonstrate its effectiveness. With a 99.89% top-2 accuracy, 99.98% top-3 accuracy, and 100% top-5 accuracy, the model ensures that the correct class is often among the top predictions, enhancing its practical applicability in scenarios where multiple potential threats need to be considered.

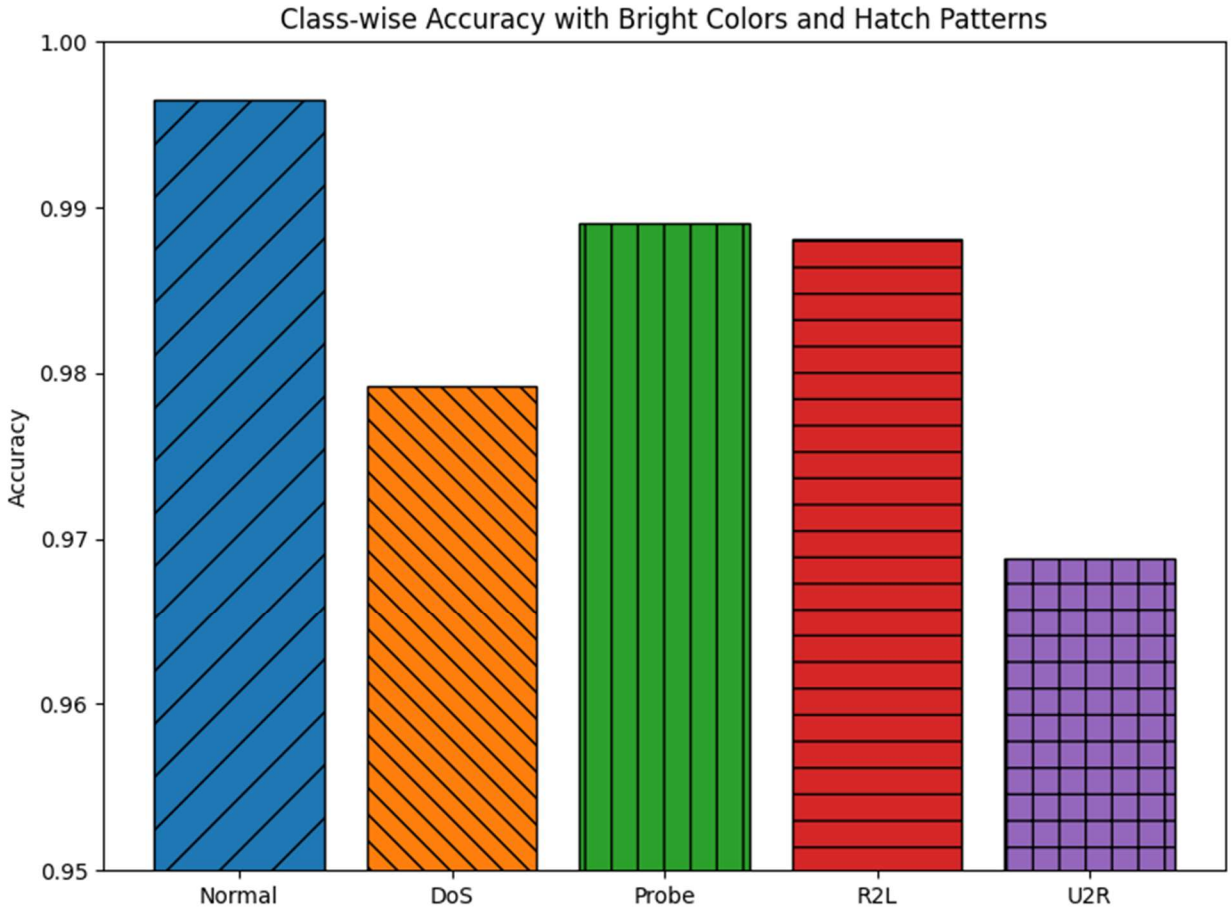

Figure 4: Performance Of The Proposed Model

Figure 5 presents the ROC and PR curves for all five classes. In the ROC curve (left), the model achieves near-perfect AUC scores for each class; Classes 0, 1, and 4 attain an AUC of 1.00, while

Journal of Theoretical and Applied Information Technology
31st July 2025. Vol.103. No.14

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Classes 2 and 3 achieve an AUC of **0.99**, indicating highly effective discrimination between classes. The ROC curve illustrates the model's performance across all classes. On the right, the Precision-Recall (PR) curve provides insights into the model's ability to handle imbalanced data and class-specific performance. The Average Precision (AP) values for Class 0, 1, and 4 are **1.00**, while Class 2 and 3 slightly trail with **0.98** and **0.97**, respectively. The nearly flat curves near the top right corner reflect high precision and recall across all classes. From this, it is concluded that the model reliably identifies relevant patterns with minimal false positives.

Figure 6 illustrates the t-SNE visualization of the test sample embeddings. This non-linear dimensionality reduction technique projects high-dimensional features into a two-dimensional space for visual inspection and analysis. Each color denotes a distinct class label. The resulting plot shows well-separated clusters, with minimal overlap between different classes. Notably, Classes 0, 1, and 3 exhibit compact and distinct clusters, demonstrating that the model has learned discriminative feature representations. Some minor overlaps exist between Class 2 and Class 4, aligning with the slightly lower PR values observed in Figure 1, but overall class separation remains strong.

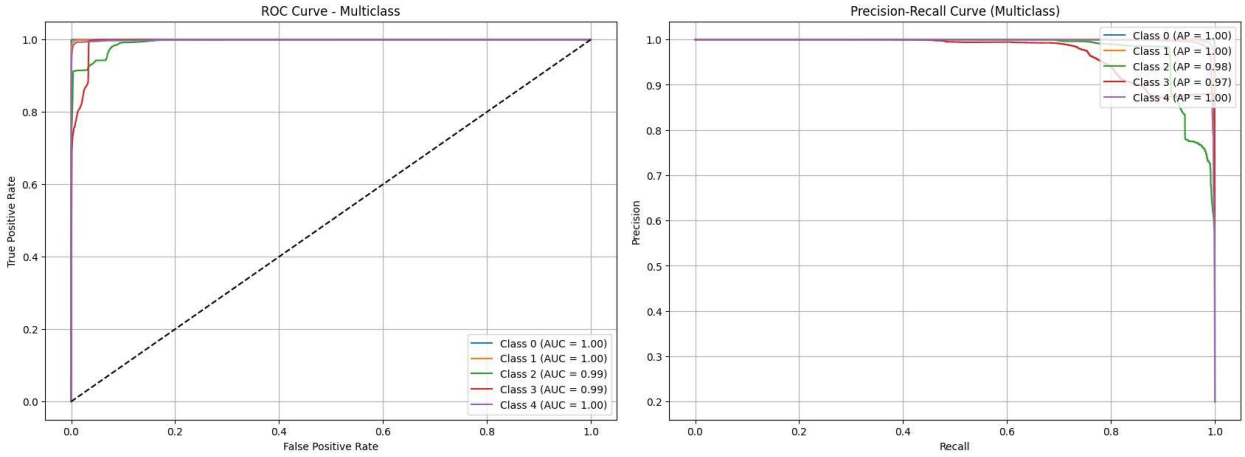

Figure 5: ROC And P-R Curve Of The Proposed Model

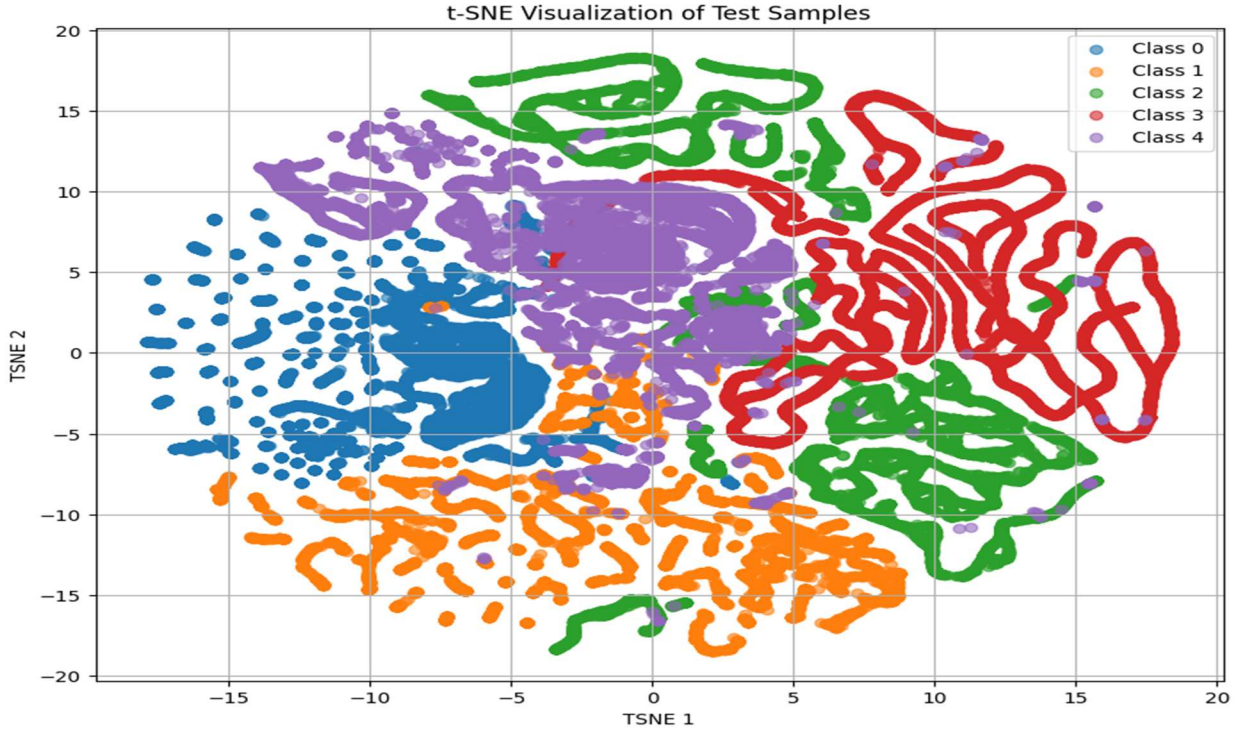

Figure 6: Class-Wise TSNE Of The Proposed Model

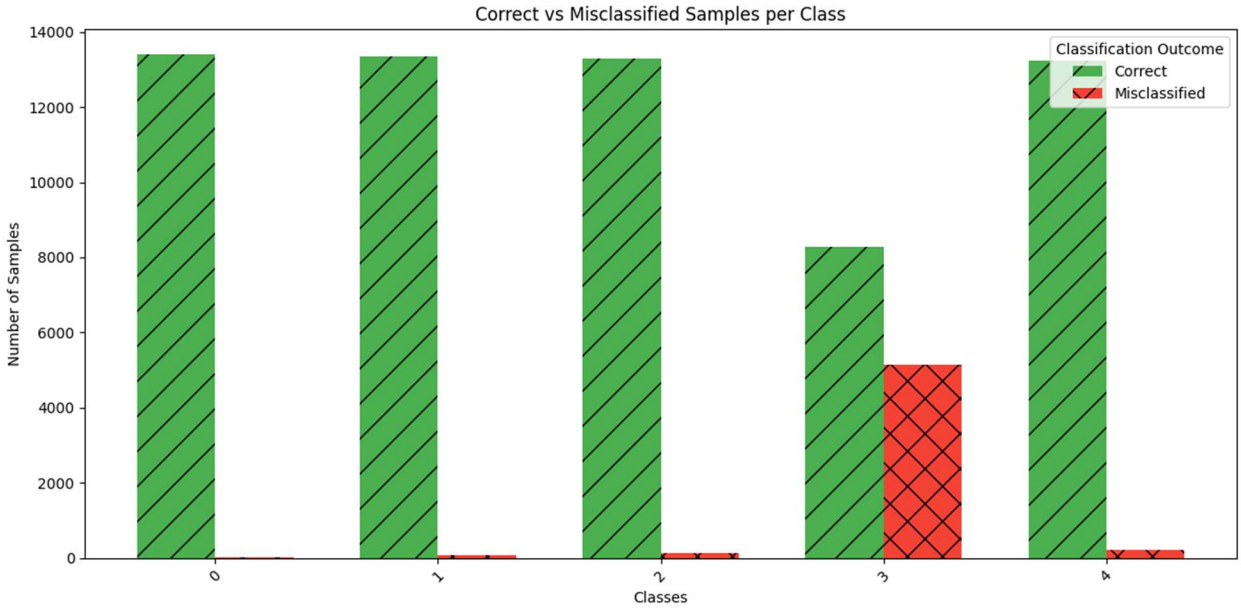

Figure 7: Correctly Classifies And Misclassifies Samples With The Proposed Model On Test Data.

From Figure 7, it is observed that the model achieves near-perfect classification accuracy for Classes 0, 1, 2, and 4, with over 13,000 instances correctly classified in each case and negligible misclassifications.

In contrast, Class 3 exhibits a noticeable drop in classification performance. Out of the total samples for Class 3, approximately 5,000 instances were misclassified, representing a significant portion of the total. This discrepancy suggests that the model struggles to distinguish Class 3 from other classes, possibly due to overlapping feature representations, class imbalance, or intrinsic similarities with other classes in the dataset.

The findings from this figure align with the slightly lower AUC (0.99) and average precision (0.97) observed for Class 3 in the ROC and PR curves.

Table 1: Comparison Of The Proposed Model With The Prescribed Model

Refence	Methodology	Dataset Used	Accuracy (%)
[3]	The hybrid CNN-RNN	NSL-KDD	81.38%
[4]	A hybrid IDS integrating anomaly detection	UNSW-NB15	85.10%
[6]	CNN, RNN, and hybrid CNN-RNN	NSL-KDD	81.38%
[8]	MHAID-IWSOA	Edge-IIoT	98.28%
[13]	Enhanced Residual-MBi-LSTM	HighD	98.01%
[16]	XGBoost for anomaly classification, SVM	NSL-KDD	94.20%
[17]	AAGCN	CMDFALL, MICA-Action3D, NTU-RGBD	88.2%

[18]	Network traffic representation, network intrusion detection	NSL-KDD	95.20%
[19]	RAE-GAMI-Net	Proof of Stake blockchain	96.20%
[20]	YAMNet IFD, CNNs	MIMII	83.55%
[21]	BiLSTM	custom cattle video	90.70%
[22]	MST3D	On the Prevention	90.94%
[24]	CNN, SVM	CASME II and SAMM	89.40%
Proposed model	LSTM + Multi-head attention	NSL-KDD	98.4%

To test the robustness of the model, we created four ablated models. We evaluated and compared their performance with that of the original model, as shown in Table 2, focusing solely on the integration of LSTM and attention mechanisms. The baseline model utilized a single-layer LSTM with 128 units, followed by a dense layer and dropout, and got an accuracy of 0.91. The second model incorporated a Multi-Head Attention (MHA) mechanism after the LSTM layer, which captures positional and sequential patterns from different parts of the input sequence, with a performance of 0.96. The third architecture, which utilized a Bidirectional LSTM to capture contextual information from both past and future states, followed by MHA, achieved an accuracy of 0.96. The final model, which combined LSTM with MHA and added two dense layers to enhance feature representation, outperformed all the other models, achieving an accuracy of 0.98. All ablated models were trained using the Adam optimizer for weight update and with early stopping based on validation loss. Performance was assessed using classification metrics on the test set. From all the ablated models, the original model performed consistently.

Table 2: Ablated Models With Hyperparameters And Accuracy

Model Name	LSTM Units	Bi-LSTM	Multi-Head Attention	Dense Layers	Dropout Rates	Global Avg Pooling	Acc
LSTM	128	No	No	1 × 128	0.2, 0.3	No	0.91
LSTM + Attention	128	No	Yes (4 heads)	1 × 128	0.2, 0.3	Yes	0.96
Bi-LSTM + Attention	64 × 2	Yes	Yes (4 heads)	1 × 128	0.2, 0.3	Yes	0.96
LSTM + Attention + Dense	128	No	Yes (4 heads)	2 × (256,128)	0.2, 0.3	Yes	0.984

5. CONCLUSION

This study presents a robust and scalable hybrid model that combines LSTM with multi-head attention to effectively capture sequential and temporal features, thereby detecting and classifying intrusion events in network traffic data. With the LSTM's temporal modeling capabilities and attention mechanisms, the model successfully captures complex patterns and highlights salient features across different attack types, achieving an accuracy of 0.984. The validated results on the NSL-KDD dataset demonstrate that the proposed model not only achieves high accuracy but also effectively handles minority classes and detects various types of adversarial attacks, which are typically challenging in intrusion detection systems (IDS). The use of the over-sampling method for data balancing, and with normalization and categorical encoding, significantly contributed to the model's generalization ability. Moreover, the model's performance is observed, including ROC, PR curves, and t-SNE plots, which reinforces the model's discriminative capability. Although the classification performance slightly declined for the R2L class, this opens avenues for further enhancements using techniques such as focal loss, advanced class-specific augmentation, or adaptive attention scaling.

REFERENCES

[1] Sharma, N., & Shambharkar, P. G. (2025). Multi-attention DeepCRNN: An efficient and explainable intrusion detection framework for Internet of Medical Things environments. *Knowledge and Information Systems*, 1-67.

[2] Kanthimathi, S., Venkatraman, S., Jayasankar, K. S., Jiljith, T. P., & Jashwanth, R. (2024). A Novel self-attention-enabled weighted ensemble-based convolutional neural network framework for distributed denial of service attack classification. *IEEE Access*.

[3] Bhattacharya, S., Khanna, A., Ganapaneni, S., & Najana, M. (2024). Attention-Based Deep Learning Frameworks for Network Intrusion Detection: An Empirical Study. *International Journal of Global Innovations and Solutions (IJGIS)*.

[4] Imrana, Y., Xiang, Y., Ali, L., Noor, A., Sarpong, K., & Abdullah, M. A. (2024). CNN-GRU-FF: a double-layer feature fusion-based network intrusion detection system using convolutional neural network and gated recurrent units. *Complex & Intelligent Systems*, *10*(3), 3353-3370.

[5] Sukanya, M., Kanimozhi, S., Kumar, G. P., & Atheeswaran, A. (2024). IOT NETWORKS'IMPROVED HYBRID INTRUSION DETECTION SYSTEM WITH ATTENTION MECHANISM. *Journal of Engineering and Technology Management*, *74*, N0.

[6] Ganapaneni, S. B. A. K. S., & Najana, M. Attention-Based Deep Learning Frameworks for Network Intrusion Detection: An Empirical Study.

[7] Cui, B., Chai, Y., Yang, Z., & Li, K. (2024). Intrusion Detection in IoT Using Deep Residual Networks with Attention Mechanisms. *Future Internet*, *16*(7), 255.

[8] Aljabri, J. (2025). Attack resilient IoT security framework using multi head attention based representation learning with improved white shark optimization algorithm. *Scientific Reports*, *15*(1), 14255.

[9] Logeswari, G., Roselind, J. D., Tamilarasi, K., & Nivethitha, V. (2025). A Comprehensive Approach to Intrusion Detection in IoT Environments Using Hybrid Feature Selection and Multi-Stage Classification Techniques. *IEEE Access*.

[10] Mao, J., Yang, X., Hu, B., Lu, Y., & Yin, G. (2025). Intrusion Detection System Based on Multi-Level Feature Extraction and Inductive Network. *Electronics*, *14*(1), 189.

[11] Oladele, D. A., & Markus, E. D. (2025). ABA-IDS: Attention-Based Autoencoder for Intrusion Detection in Assistive mobility robotic Network. *IEEE Access*.

[12] Vadisetty, R. (2024, March). Adaptive Machine Learning-Based Intrusion Detection Systems for IoT Era. In *International Ethical Hacking Conference* (pp. 251-273). Singapore: Springer Nature Singapore.

[13] Wu, Z., Liang, K., Liu, D., & Zhao, Z. (2022). Driver lane change intention recognition based on Attention Enhanced Residual-MBi-LSTM network. *IEEE Access*, *10*, 58050-58061.

[14] Awan, M. J., Masood, O. A., Mohammed, M. A., Yasin, A., Zain, A. M., Damaševičius, R., & Abdulkareem, K. H. (2021). Image-based malware classification

using VGG19 network and spatial convolutional attention. *Electronics*, *10*(19), 2444.

[15] Liao, N., & Guan, J. (2024). Multi-scale Convolutional Feature Fusion Network Based on Attention Mechanism for IoT Traffic Classification. *International Journal of Computational Intelligence Systems*, *17*(1), 36.

[16] Immadisetty, A. Machine Learning for Real-Time Anomaly Detection.

[17] Pham, D. T., Pham, Q. T., Le, T. L., & Vu, H. (2021). An efficient feature fusion of graph convolutional networks and its application for real-time traffic control gestures recognition. *IEEE Access*, *9*, 121930-121943.

[18] Zhang, J., Shi, Z., Wu, H., & Xing, M. (2022, November). A Novel Self-supervised Few-shot Network Intrusion Detection Method. In *International Conference on Wireless Algorithms, Systems, and Applications* (pp. 513-525). Cham: Springer Nature Switzerland.

[19] Chithanuru, V., & Ramaiah, M. (2025). An Efficient approach based on RAE-GAMI-NET for Long Range attack Detection on Blockchain. *IEEE Access*.

[20] Alharbi, F., Luo, S., Zhao, S., Yang, G., Wheeler, C., & Chen, Z. (2024). Belt Conveyor Idlers Fault Detection Using Acoustic Analysis and Deep Learning Algorithm with the YAMNet Pretrained Network. *IEEE Sensors Journal*.

[21] Qiao, Y., Clark, C., Lomax, S., Kong, H., Su, D., & Sukkarieh, S. (2021). Automated individual cattle identification using video data: A unified deep learning architecture approach. *Frontiers in Animal Science*, *2*, 759147.

[22] Chao, X., Qi, X., Ding, R., & Ji, G. (2025). Vehicle lane change behavior recognition based on multi-scale three-stream 3D ResNets. *Multimedia Systems*, *31*(2), 129.

[23] Jiang, Z., Wang, C., Li, J., Zhao, M., & Yang, Q. (2024). Self-supervised memory-guided and attention feature fusion for video anomaly detection. *Journal of Electronic Imaging*, *33*(6), 063035-063035.

[24] Malik, P., Harsola, A., Sharma, S., Chouhan, S., Rao, A. S., & Bhadouriya, A. (2025). Micro-Expression Recognition for Lie Detection Using Image Processing. *Image*, *13*, 2.

[25] Abd El-Nabi, S., Ibrahim, A. F., El-Rabaie, E. S. M., Hassan, O. F., Soliman, N. F., Ramadan, K. F., & El-Shafai, W. (2025). Driver Drowsiness Detection Using Swin Transformer and Diffusion Models for Robust Image Denoising. *IEEE Access*.