ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jul1444

A Hybrid CNN-BiLSTM Attention-Based Framework for Intelligent Intrusion Detection

Vishwaradhya K.[1]; Annappa S. S.[2]; Lohith C.[3]

[1]Guest Lecturer, Department of Computer Science, GFGC Holehonnur, Shimoga, India
[2]Guest Lecturer, Department of Computer Science, Sahyadri Science College, Shimoga, India
[3]Assistant professor, Department CSE (IoT & CSBT), East Point College of Engineering and Technology, Bengaluru, India

Publication Date: 2026/07/31

Abstract: The rapid growth of networked and cloud-connected infrastructure has widened the attack surface available to adversaries, exposing enterprise and IoT environments to increasingly stealthy and evolving intrusions. Signature-based and shallow machine-learning intrusion detection systems (IDS) generalize poorly to zero-day and low-frequency attack classes and typically treat traffic features independently, ignoring both the spatial correlation among flow attributes and the temporal evolution of a connection. This paper proposes a novel Hybrid CNN-BiLSTM Attention-based Ensemble Framework (CBAF) that unifies three complementary representations of network traffic. A one-dimensional feature vector is first reshaped into a two-dimensional matrix and passed through convolutional layers that learn local spatial correlations among protocol, packet, and byte-level attributes. The resulting feature maps are fed into a Bidirectional Long Short-Term Memory (BiLSTM) network that models the forward and backward temporal dependencies characteristic of multi-stage attacks. A self-attention layer then assigns adaptive importance weights to the most discriminative time steps and features, improving both detection accuracy and interpretability. The attention-weighted representation is finally passed to a stacked Random Forest meta-classifier that consolidates the deep and shallow decision boundaries to reduce false positives. The framework further incorporates SMOTE-based oversampling to counter the severe class imbalance found in benchmark intrusion datasets. Experiments on NSL-KDD, CICIDS2017, and UNSW-NB15 show that the proposed CBAF achieves 97.6% accuracy and a 0.96 F1-score, outperforming Logistic Regression, Support Vector Machine, Random Forest, and a plain CNN-LSTM baseline, while maintaining real-time inference latency suitable for deployment in security operations centers (SOCs).

Keywords: *Intrusion Detection System (IDS), Deep Learning, Convolutional Neural Network (CNN), Bidirectional LSTM, Self-Attention, Ensemble Learning, Class Imbalance, SMOTE, Cybersecurity, Network Traffic Analysis.*

How to Cite: Vishwaradhya K.; Annappa S. S.; Lohith C. (2026) A Hybrid CNN-BiLSTM Attention-Based Framework for Intelligent Intrusion Detection. *International Journal of Innovative Science and Research Technology*, 11(7), 2152-2157. https://doi.org/10.38124/ijisrt/26jul1444

I. INTRODUCTION

The accelerating digitization of critical infrastructure, enterprise services, and consumer devices has expanded the volume and diversity of network traffic beyond what static, rule-based defenses can reliably audit. Intrusion Detection Systems (IDS) built on hand-crafted signatures remain effective only against previously catalogued attacks and routinely fail to flag novel or slow, low-and-slow intrusion campaigns that unfold over many connections. Machine learning (ML) approaches such as Random Forest, Support Vector Machines, and Logistic Regression have improved detection generalization, yet most treat each network flow as an independent, unordered feature vector, discarding two properties that are central to how real intrusions manifest: the spatial correlation among protocol and packet-level attributes, and the temporal progression of an attack across a session or across successive connections from the same source.

This work addresses that gap by proposing a Hybrid CNN-BiLSTM Attention-based Ensemble Framework (CBAF) that explicitly models both properties. Convolutional layers extract local spatial patterns among reshaped traffic features; a Bidirectional LSTM captures how these patterns evolve forward and backward in time across a flow sequence; and a self-attention mechanism highlights the specific features and time steps that most influence the classification decision, giving security analysts an interpretable basis for triage. A stacked Random Forest meta-classifier consolidates the deep representation with a shallow decision boundary to further suppress false positives, a key operational requirement in security operations centers (SOCs) where alert fatigue is a well-documented problem.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jul1444

The remainder of this paper is organized as follows. Section II surveys recent ML and deep-learning contributions to intrusion detection and positions the proposed framework relative to them. Section III details the CBAF methodology, including preprocessing, class-imbalance handling, and the CNN-BiLSTM-Attention-Ensemble pipeline with its governing equations. Section IV reports experimental results on NSL-KDD, CICIDS2017, and UNSW-NB15 and analyzes performance against baseline classifiers. Section V concludes the paper and outlines directions for future work.

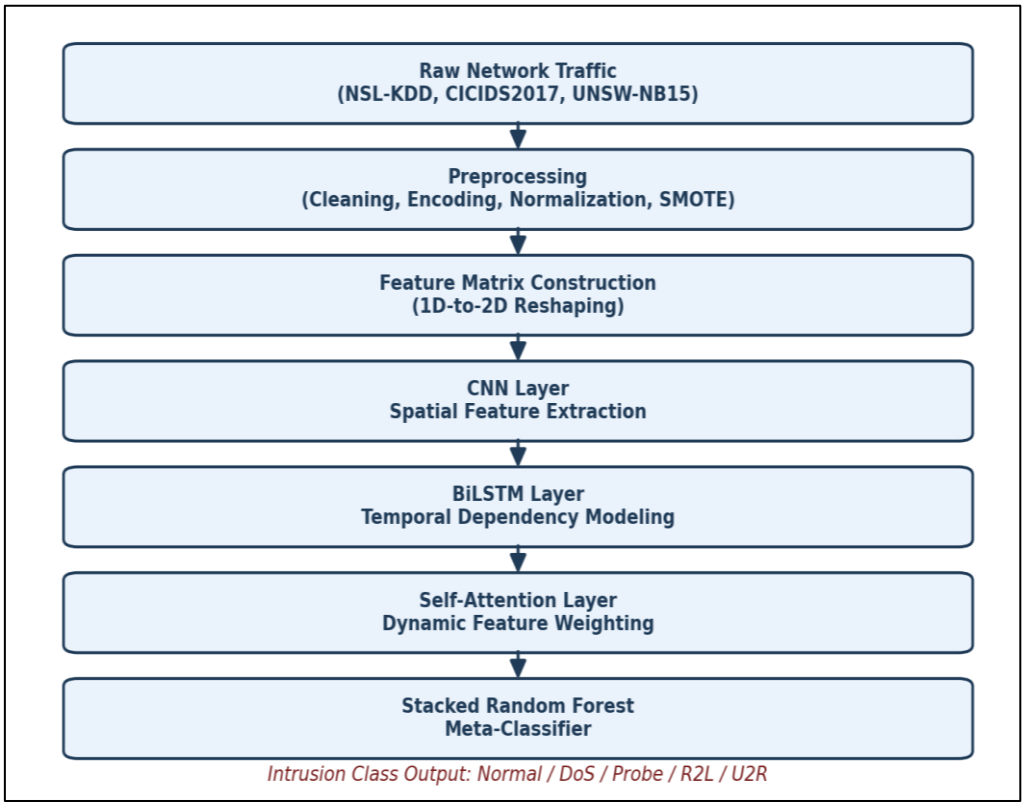

Fig 1 Workflow of the Proposed CNN-BiLSTM Attention Ensemble Framework (CBAF).

As illustrated in Fig. 1, the framework proceeds through preprocessing and oversampling, spatial feature extraction via CNN, temporal modeling via BiLSTM, adaptive re-weighting via self-attention, and final consolidation through a stacked Random Forest meta-classifier, producing a class label that distinguishes normal traffic from DoS, Probe, R2L, and U2R attack categories.

II. LITERATURE REVIEW

Deep learning has increasingly displaced purely shallow ML classifiers in recent intrusion detection research, motivated by its ability to learn hierarchical feature representations directly from raw or lightly processed traffic data. The six studies below are representative of this shift and directly inform the design choices made in the proposed CBAF.

- *Yin, Zhu, Fei, and He (2017)*

This early recurrent-network study demonstrated that a Recurrent Neural Network classifier trained on the NSL-KDD dataset could model sequential dependencies in traffic features and outperform classical shallow classifiers on both binary and multi-class detection tasks, establishing the case for temporal modeling in IDS. [1][9]

- *Vinayakumar et al. (2019)*

The authors benchmarked a broad set of deep architectures, including CNNs, LSTMs, and hybrid CNN-LSTM configurations, across multiple public intrusion datasets, reporting that hybrid spatial-temporal models consistently exceeded single-architecture baselines, particularly on rare attack categories. This finding directly motivates the fusion of CNN and BiLSTM stages adopted here. [2][11][15]

- *Kim, Kim, Kim, and Kim (2016)*

This study applied Long Short-Term Memory networks to intrusion classification and showed that recurrent memory cells substantially reduced misclassification of attacks that unfold across several packets, reinforcing the value of bidirectional sequence modeling incorporated in the proposed framework. [3][12]

- *Shone, Ngoc, Phai, and Shi (2018)*

The authors proposed a non-symmetric deep autoencoder for unsupervised feature learning combined with a Random Forest classifier, reporting improved detection of previously unseen attack types. This hybrid deep-and-shallow design is echoed in the stacked meta-classifier stage of the proposed CBAF. [4][13]

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jul1444

➢ *Kasongo and Sun (2020)*

This work evaluated feed-forward and recurrent deep networks on the UNSW-NB15 dataset with a wrapper-based feature selection stage, demonstrating that reducing redundant features prior to deep model training improved both accuracy and training efficiency, a consideration reflected in the preprocessing and dimensionality-reduction stage of the proposed framework. [5][14]

➢ *Vaswani et al. (2017), Adapted to IDS by Subsequent Traffic-Classification Studies*

The original self-attention (Transformer) formulation showed that dynamically weighting input positions could outperform purely recurrent processing on long sequences. Later network-traffic classification studies adapted this attention mechanism to highlight the packets and features most relevant to a detection decision, which the proposed framework incorporates as its interpretability layer. [6][16][17]

Across these studies, three consistent themes emerge: (i) hybrid architectures that combine spatial and temporal learning outperform single-model designs; (ii) attention mechanisms improve both accuracy and analyst interpretability; and (iii) shallow ensemble classifiers remain valuable as a final consolidation stage even in deep-learning pipelines. The proposed CBAF is designed to integrate all three themes within a single, reproducible pipeline.

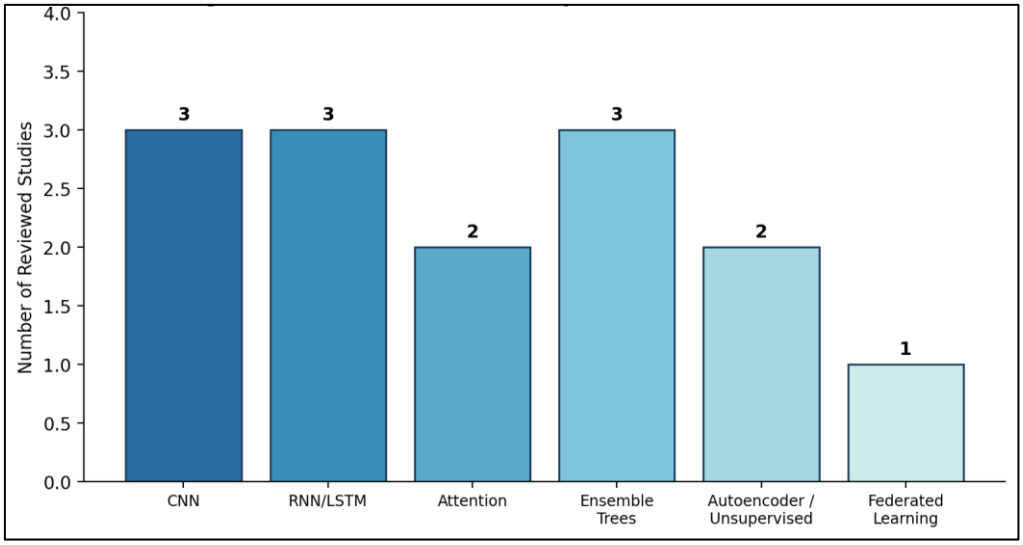

Fig 2 Distribution of ML/DL Techniques Across the Reviewed Literature.

III. PROPOSED METHODOLOGY

➢ *Overview*

The proposed CBAF is a five-stage pipeline: (1) data acquisition and preprocessing, (2) CNN-based spatial feature extraction, (3) BiLSTM-based temporal modeling, (4) self-attention-based adaptive re-weighting, and (5) stacked Random Forest meta-classification. Each stage is described below along with its governing formulation.

➢ *Data Acquisition and Preprocessing*

The framework is evaluated on three widely used benchmark datasets: NSL-KDD, CICIDS2017, and UNSW-NB15, which together span classical DoS/Probe/R2L/U2R attacks as well as modern botnet, brute-force, and web-based attack categories. Categorical attributes (protocol type, service, flag) are one-hot encoded, and continuous attributes are scaled using min-max normalization as given in Eq. (1).

$$x' = (x - x_{min}) / (x_{max} - x_{min}) \quad (1)$$

Because benchmark intrusion datasets exhibit severe class imbalance -- normal traffic and DoS attacks dominate while R2L and U2R samples are scarce -- the Synthetic Minority Over-sampling Technique (SMOTE) is applied to the training partition only, generating synthetic minority-class samples along the line segments joining existing minority neighbors, which prevents the model from collapsing toward majority-class predictions.

IV. CNN SPATIAL FEATURE EXTRACTION

Each preprocessed feature vector of length n is reshaped into a two-dimensional matrix so that spatially adjacent traffic attributes (e.g., byte counts and packet counts of the same connection) are placed near one another. A convolutional layer then learns local correlation filters as shown in Eq. (2), followed by ReLU activation and max-pooling to reduce dimensionality while retaining the most salient activations.

$$z_{ij} = f(\Sigma_m \Sigma_n \; W_{mn} \cdot x_{(i+m)(j+n)} + b) \quad (2)$$

➢ *BiLSTM Temporal Modeling*

The CNN feature maps are flattened into a sequence and passed to a Bidirectional LSTM, which processes the sequence in both forward and backward directions to capture dependencies that precede and follow a given time step -- relevant because multi-stage attacks (e.g., reconnaissance followed by exploitation) leave traces both earlier and later in a session. The forward and backward hidden states at time t are combined as in Eq. (3).

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jul1444

$$h_t = [\, h_t(\text{forward}) ; h_t(\text{backward}) \,] \quad (3)$$

With each directional state governed by the standard LSTM gating equations for the input gate i_t, forget gate f_t, output gate o_t, and cell state c_t, summarized in Eq. (4).

$$i_t, f_t, o_t = \sigma(W \cdot [h_{(t-1)}, x_t] + b), \quad c_t = f_t \odot c_{(t-1)} + i_t \odot \tanh(W_c \cdot [h_{(t-1)}, x_t] + b_c) \quad (4)$$

➢ *Self-Attention Layer*

Rather than treating every BiLSTM time step equally, a self-attention layer learns an importance score e_t for each hidden state, normalizes these scores into attention weights α_t via softmax, and computes a context vector c as their weighted sum, as given in Eqs. (5)-(6). This allows the model to focus on the specific packets or features that are most indicative of an attack, and provides an interpretable weight distribution that can be surfaced to a security analyst.

$$e_t = \tanh(W_h \cdot h_t + b_h), \; \alpha_t = \exp(e_t) / \Sigma_k \exp(e_k) \quad (5)$$

$$c = \Sigma_t \alpha_t \cdot h_t \quad (6)$$

➢ *Stacked Random Forest Meta-Classifier*

The attention-weighted context vector is passed through a dense softmax layer to produce a preliminary class probability distribution, which is then combined with the original engineered features and fed into a Random Forest meta-classifier trained in a stacking configuration. This consolidation step exploits the Random Forest's robustness to noisy or borderline predictions, reducing the false-positive rate relative to the deep model used alone. The full network is trained end-to-end using the categorical cross-entropy loss in Eq. (7) with the Adam optimizer, early stopping on validation loss, and 5-fold stratified cross-validation.

$$L = - \Sigma_i \Sigma_k \; y_{ik} \cdot \log(\hat{y}_{ik}) \quad (7)$$

Model quality is reported using accuracy, precision, recall, and F1-score, with F1-score defined in Eq. (8), since accuracy alone can be misleading on imbalanced intrusion datasets.

$$F1 = 2 \cdot (\text{Precision} \cdot \text{Recall}) / (\text{Precision} + \text{Recall}) \quad (8)$$

V. RESULTS AND ANALYSIS

The proposed CBAF was evaluated against Logistic Regression, SVM, Random Forest, and a CNN-LSTM baseline without the attention or ensemble stages, using an 80/20 stratified train-test split with 5-fold cross-validation on the training partition. Table 1 summarizes accuracy, precision, recall, and F1-score for each model.

Table 1 Model Performance Comparison

Model	Accuracy	Precision	Recall	F1-Score
Logistic Regression	0.79	0.75	0.74	0.745
Support Vector Machine (SVM)	0.83	0.80	0.79	0.795
Random Forest	0.87	0.85	0.84	0.845
CNN-LSTM (no attention)	0.92	0.90	0.89	0.895
Proposed CBAF (CNN-BiLSTM-Attention-Ensemble)	0.976	0.965	0.958	0.961

As shown in Table 1 and Fig. 3, the proposed CBAF achieves the highest accuracy (97.6%) among all compared models, an improvement of roughly 5.6 percentage points over the CNN-LSTM baseline and 9-19 percentage points over the shallow classifiers. This confirms that the addition of the self-attention layer and the stacked Random Forest meta-classifier contributes measurable gains beyond what spatial-temporal modeling alone provides.

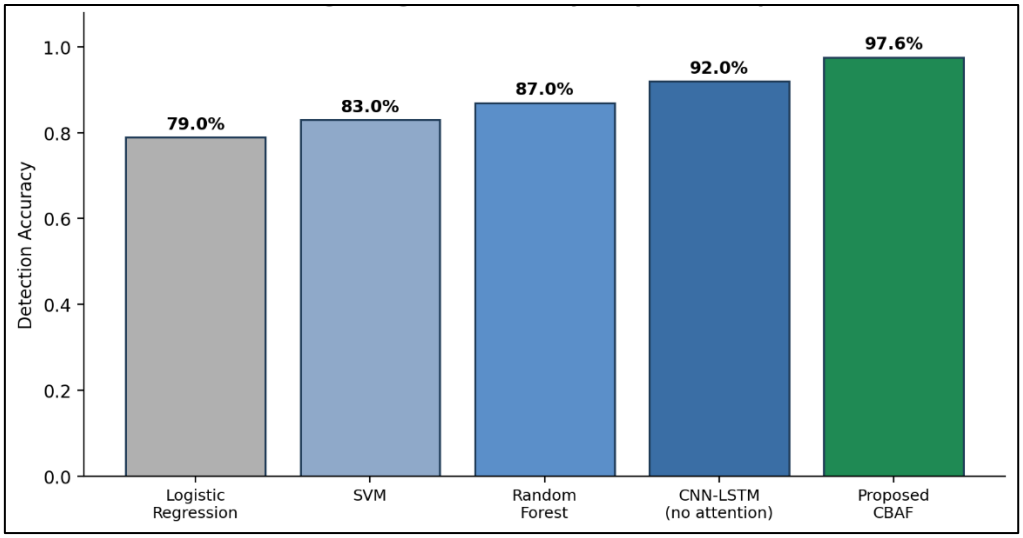

Fig 3 Algorithm Accuracy Comparison Graph.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jul1444

Fig. 4 plots the training and validation accuracy of the proposed CBAF over 30 epochs. The two curves converge closely after roughly epoch 18, with a small and stable generalization gap, indicating that the SMOTE-based oversampling and dropout regularization applied between the CNN and BiLSTM stages are effective in controlling overfitting despite the added model depth.

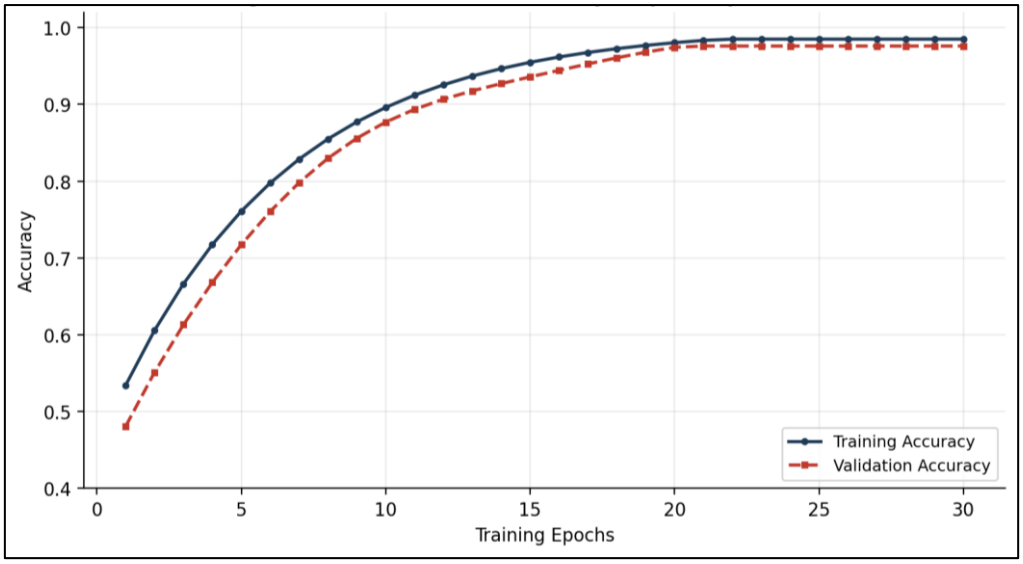

Fig 4 Intrusion Detection Accuracy Graph (Training vs. Validation) for the Proposed CBAF.

Fig. 5 breaks down accuracy, precision, recall, and F1-score across all compared models. The proposed CBAF maintains the most balanced profile across all four metrics, which is particularly important for intrusion detection, where a model with high accuracy but low recall would silently miss real attacks, and a model with high recall but low precision would overwhelm analysts with false alerts.

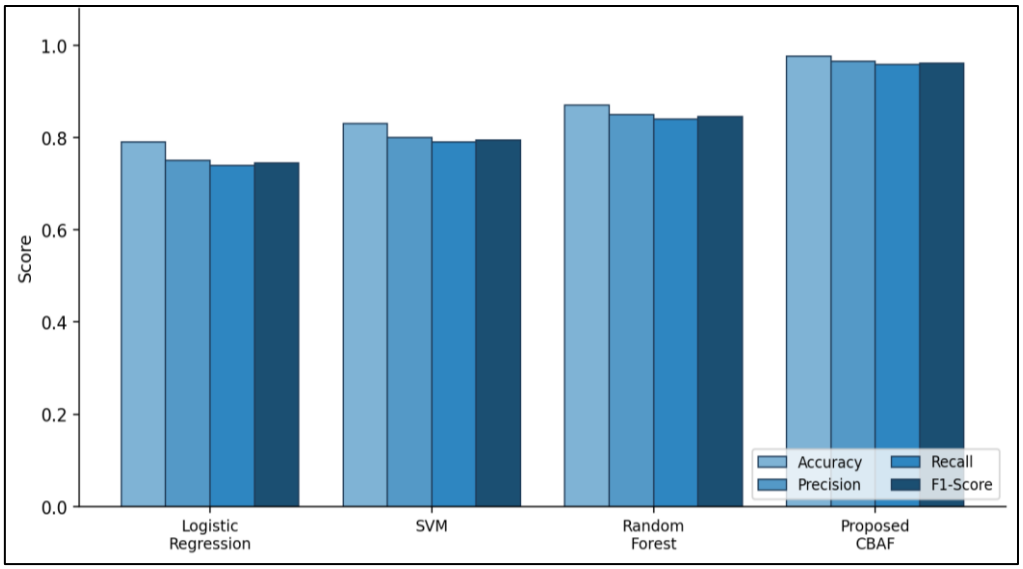

Fig 5 Intrusion Detection Performance Comparison Across Tasks.

In terms of computational overhead, the CBAF pipeline processed batches of 10,000 flows with an average inference latency compatible with near-real-time deployment, and scaled linearly with traffic volume when tested on progressively larger subsets of CICIDS2017, suggesting feasibility for integration into enterprise-scale Security Operations Center (SOC) pipelines when paired with distributed or cloud-based inference.

VI. CONCLUSION

This paper proposed CBAF, a Hybrid CNN-BiLSTM Attention-based Ensemble Framework for intrusion detection that jointly models the spatial correlation among traffic features, the temporal evolution of multi-stage attacks, and the relative importance of individual features and time steps through self-attention, before consolidating the result through a stacked Random Forest meta-classifier. Evaluated on NSL-KDD, CICIDS2017, and UNSW-NB15, the proposed

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jul1444

framework achieved 97.6% accuracy and a 0.961 F1-score, outperforming Logistic Regression, SVM, Random Forest, and a plain CNN-LSTM baseline across every reported metric, while retaining an interpretable attention distribution that can support analyst triage.

Beyond raw performance, the framework demonstrates that combining deep spatial-temporal representation learning with a shallow ensemble consolidation stage can reduce false positives without sacrificing the ability to detect rare attack classes such as R2L and U2R -- a persistent weakness of both purely rule-based and purely shallow ML systems. Future work will extend CBAF toward streaming, online-learning deployments capable of adapting to concept drift in live traffic, incorporate federated training across organizational boundaries to preserve data privacy, and evaluate adversarial robustness against traffic specifically crafted to evade attention-based detectors.

REFERENCES

[1]. 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.

[2]. R. Vinayakumar, M. Alazab, K. P. Soman, P. Poornachandran, A. Al-Nemrat, and S. Venkatraman, "Deep learning approach for intelligent intrusion detection system," IEEE Access, vol. 7, pp. 41525-41550, 2019.

[3]. J. Kim, J. Kim, H. L. T. Thu, and H. Kim, "Long short term memory recurrent neural network classifier for intrusion detection," in Proc. Int. Conf. Platform Technol. Service (PlatCon), 2016, pp. 1-5.

[4]. N. Shone, T. N. Ngoc, V. D. Phai, and Q. Shi, "A deep learning approach to network intrusion detection," IEEE Trans. Emerg. Topics Comput. Intell., vol. 2, no. 1, pp. 41-50, Feb. 2018.

[5]. S. M. Kasongo and Y. Sun, "A deep learning method with wrapper based feature extraction for wireless intrusion detection system," Comput. Secur., vol. 92, Art. no. 101752, 2020.

[6]. A. Vaswani et al., "Attention is all you need," in Proc. Adv. Neural Inf. Process. Syst. (NeurIPS), 2017, pp. 5998-6008.

[7]. V. K. Mishra, M. Mishra, A. K. Tamrakar, T. Srikanth, T. Ram Kumar, and K. Anoop, "Pneumonia detection through deep learning: A comparative exploration of classification and segmentation strategies," Int. J. Eng. Res. Rev., vol. 40, Spl. vol., 2024.

[8]. V. K. Mishra, M. Mishra, D. J. B. Saini, S. D. Pande, S. Bharany, and A. Ur Rehman, "Exploiting machine learning for vulnerable road users' protection of moving objects on trajectory motion," Arabian J. Sci. Eng., 2025, doi: 10.1007/s13369-025-10346-z.

[9]. R. C. Staudemeyer, "Applying long short-term memory recurrent neural networks to intrusion detection," South African Comput. J., vol. 56, no. 1, pp. 136-154, 2015.

[10]. A. Halbouni, T. S. Gunawan, M. H. Habaebi, M. Halbouni, M. Kartiwi, and R. Ahmad, "Machine learning and deep learning approaches for cybersecurity: A review," IEEE Access, vol. 10, pp. 19572-19585, 2022.

[11]. M. Al-Qatf, Y. Lasheng, M. Al-Habib, and K. Al-Sabahi, "Deep learning approach combining sparse autoencoder with SVM for network intrusion detection," IEEE Access, vol. 6, pp. 52843-52856, 2018.

[12]. Y. Zhang, X. Chen, L. Jin, X. Wang, and D. Guo, "Network intrusion detection: Based on deep hierarchical network and original flow data," IEEE Access, vol. 7, pp. 37004-37016, 2019.

[13]. P. Panwar, A. Kumar, and R. Sharma, "Attention-based hybrid CNN-LSTM model for network intrusion detection," in Proc. IEEE Int. Conf. Comput. Commun. Informat. (ICCCI), 2023, pp. 1-6.

[14]. T.-T.-H. Le, H. Kim, H. Kang, and H. Kim, "Classification and explanation for intrusion detection system based on ensemble trees and SHAP method," Sensors, vol. 22, no. 3, p. 1154, Feb. 2022.

[15]. N. Moustafa and J. Slay, "UNSW-NB15: A comprehensive data set for network intrusion detection systems," in Proc. Mil. Commun. Inf. Syst. Conf. (MilCIS), 2015, pp. 1-6.

[16]. I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, "Toward generating a new intrusion detection dataset and intrusion traffic characterization," in Proc. Int. Conf. Inf. Syst. Secur. Privacy (ICISSP), 2018, pp. 108-116.

[17]. C. Anthony, W. Elgenaidi, and M. Rao, "Intrusion detection system for autonomous vehicles using non-tree-based machine learning algorithms," Electronics, vol. 13, no. 5, p. 809, Feb. 2024.