“Toward Stronger Intrusion Detection: Behavioral Biometrics and Network Data Fusion with Deep Learning”

Aporbo Ghosh[1] **Sabrina Akter Sabina**[1] **Myesha Tabassum**[1]

Electrical and Computer Engineering, North South University, Dhaka, Bangladesh
aporbo.ghosh@northsouth.edu[1]
sabrina.sabina@northsouth.edu[1]
myesha.tabassum@northsouth.edu[1]

Abstract:
In an era of increasingly complex cyber threats, relying on a single data modality for intrusion detection often leads to suboptimal detection accuracy and high false positive rates. This paper proposes a novel multi-modal intrusion detection framework that fuses user behavioral patterns with network traffic analysis to provide robust and accurate anomaly detection. The behavioral component leverages the ISOT Web Interactions dataset and employs three deep learning architectures—standard LSTM autoencoder, deep LSTM autoencoder, and attention-based LSTM autoencoder—to model normal user behavior and detect deviations using reconstruction error. Parallelly, a 1D Convolutional Neural Network (CNN) is trained on the NSL-KDD dataset to classify network-based intrusions. To unify the outputs of these heterogeneous models, we introduce a multi-level fusion strategy consisting of a Meta Neural Network, a Super Meta Neural Network, a weighted voting scheme, and a LightGBM-based ensemble. Experimental results demonstrate that while the individual behavioral models achieve perfect precision with a recall of 0.8990, and the CNN attains a precision of 0.9681 with moderate recall, the fusion models significantly enhance overall performance. Notably, the LightGBM fusion approach achieves the best trade-off, with an F1-score of 0.9032 and accuracy of 88.96% on a large-scale evaluation. This study highlights the strength of integrating behavioral and network data for scalable, high-fidelity intrusion detection in real-world environments.

1. Introduction

Intrusion Detection Systems (IDS) are a cornerstone of modern cybersecurity infrastructure, designed to monitor, detect, and respond to malicious activities targeting organizational networks and systems. They play a pivotal role in preserving data confidentiality, ensuring system availability, and maintaining the integrity of digital operations. Traditionally, IDS architectures have relied on either network traffic analysis or user behavior profiling as standalone detection mechanisms. While effective to a degree, such unimodal approaches are increasingly inadequate against advanced persistent threats and sophisticated attack strategies.

Contemporary cyberattacks often employ stealthy techniques such as behavioral mimicry, zero-day exploits, and multi-vector intrusions that can evade detection by exploiting the blind spots inherent in single-modality systems. For instance, a network-based IDS may overlook an attack masked within legitimate traffic patterns, while a behavior-only system might fail to detect subtle network anomalies indicative of a breach. This highlights the urgent need for more comprehensive and adaptive intrusion detection strategies.

To address these limitations, this research proposes a novel **multi-modal intrusion detection framework** that combines both user behavioral analytics and network traffic inspection to deliver a more robust and context-aware detection capability. Specifically, the behavioral component utilizes the ISOT Web Interactions dataset, capturing session-level data such as keystroke dynamics, mouse trajectories, and site navigation behavior. Simultaneously, the network-based analysis leverages the NSL-KDD dataset—a widely used benchmark in intrusion detection research—to model traffic-level anomalies.

The fusion of these modalities is facilitated through advanced deep learning architectures. Three variants of LSTM-based autoencoders (standard, deep, and attention-based) are employed for unsupervised behavioral anomaly

detection, while a 1D Convolutional Neural Network (CNN) is trained for supervised classification of network intrusions. To effectively integrate the predictive strengths of these models, a layered fusion strategy is proposed, including score-level meta learning, stacked ensemble networks, and gradient boosting (LightGBM).
This comprehensive approach enables the system to cross-validate alerts from multiple perspectives, thereby improving detection accuracy, reducing false positives, and enhancing resilience against complex attack vectors. Experimental results confirm that the fusion-based framework significantly outperforms individual models, demonstrating the potential of multi-modal IDS as a scalable and intelligent defense mechanism in evolving cyber threat environments.

2. LITERATURE REVIEW

Authentication and Behavioral Biometrics Promising options for consistent and intuitive authentication are provided by behavioral biometrics. A thorough analysis of behavioral biometric modalities, including keystroke dynamics, mouse movements, and gait patterns, was presented by Ahmed et al. [1]. They emphasized issues including the difficulty of spoof detection and the substantial intra-user variability. Bleha et al. [2] examined timing features and classification methods with a particular focus on keystroke dynamics, observing notable variations in accuracy across various settings and user behaviors.

IoT and Conventional Network Intrusion Detection For the protection of both conventional and Internet of Things networks, intrusion detection systems (IDS) are essential. In their analysis of IDS methods for IoT contexts, Hindy et al. [3] noted the drawbacks caused by a variety of communication protocols and limited device resources. A thorough analysis of deep learning-based intrusion detection systems was provided by Kim et al. [4], who evaluated CNNs, RNNs, and hybrid architectures. Wang et al. [18] highlighted the benefits of hybrid and unsupervised models and provided a broad classification of machine learning techniques for anomaly identification. In their surveys of the expanding corpus of deep learning applications in IDS, Zuech et al. [20] and Almiani et al. [21] assessed the computing demands, model complexity, and detection efficacy. Buczak and Guven [22] highlighted a trade-off between interpretability and detection capability in their comparison of contemporary machine learning algorithms with conventional data mining techniques.

IDS Techniques Based on Deep Learning Performance gains have been notable when deep learning methods were incorporated into IDS frameworks. A CNN-based intrusion detection system for wireless networks was proposed by Tang et al. [5] and demonstrated good accuracy in real-time situations. Similar to this, Vinayakumar et al. [6] looked at deep feedforward networks and LSTM for intrusion detection, showing better results than traditional methods. A hybrid approach that combines fuzzy logic and deep neural networks was presented by Bostani et al. [11] to increase system adaptability in the face of uncertainty. Using machine learning approaches, cloud-based detection systems, like the one suggested by Agrawal and Thakkar [10], showed efficacy and scalability. In order to facilitate proactive detection, Lopez-Martin et al. [13] used deep learning for network traffic classification. Javaid et al. [12] focused on knowledge representation and used deep learning to extract cyber threat intelligence. While Hoonselaar and Šnajder [24] thoroughly assessed several deep learning models on standard datasets, talking about their detection performance and training complexities, Thamilarasan and Kuppusamy [17] implemented a deep neural network with competitive accuracy for IoT-specific deployments.

Explainable and Adversarial AI in Security Deep learning's expanding application in security has sparked questions about its transparency and resilience. Adversarial attacks that take use of deep learning-based IDS flaws were examined by Shafi et al. [16] and Xia et al. [19], who found that these models are easily circumvented. Explainable artificial intelligence (XAI) in cybersecurity was investigated by Sharafaldin et al. [9] in order to overcome these constraints, encouraging model transparency and human interpretability for improved decision-making.

Different Frameworks for Detection Other strategies have been put forth to increase the privacy and flexibility of detection. In order to encourage decentralized learning and minimize data exposure, Khan et al. [7] created an industrial IoT IDS architecture based on federated learning. Liu et al. [8] presented graph-based anomaly detection, which uses topological patterns in network traffic to identify harmful actions. Self-organizing maps (SOM), which provide both detection and interpretability, were used by Garcia and Plattner [23] to display and identify abnormalities in network data. In their review of attack vectors aimed at network infrastructure, Marchal et al. [14] underlined the necessity of more robust and context-aware detection systems.

Benchmark Datasets for Research on IDS To train and assess IDS models, high-quality datasets are necessary. The UNSW-NB15 dataset was presented by Moustafa and Slay [15] and includes realistic network behavior and contemporary attack methods. A meta-review of existing IDS datasets was carried out by Ring et al. [25], who compared the datasets diversity, scope, and suitability for use in modern network contexts.

3. STATE OF THE ART PROBLEM

Conventional Intrusion Detection Systems (IDS) typically operate using either network-based or host-based data. Network-based IDS leverage traffic features and have shown strong accuracy using models like SVMs and 1D-CNNs on datasets such as NSL-KDD. However, they often suffer from low recall and fail to detect stealthy or zero-day attacks.

Host-based IDS, on the other hand, utilize behavioral biometrics (e.g., keystrokes, mouse dynamics) to identify anomalies. While LSTM-based autoencoders have proven effective in modeling sequential behavior, they are limited by small-scale datasets and lack of context from network activity.

Recent works have explored multi-modal fusion, combining behavioral and network features. Yet, most rely on basic techniques like score averaging or fixed-weight voting, which lack flexibility and learning capability. Few models incorporate deep fusion layers or attention mechanisms to fully exploit cross-modal patterns.

To bridge these gaps, our work introduces a comprehensive fusion framework integrating deep behavioral modeling, 1D-CNN network classification, and multi-level fusion using trainable meta models and LightGBM ensembles—offering improved detection performance and adaptability.

4.METHODOLOGY

This paper suggests a hybrid intrusion detection system (IDS) that uses ensemble fusion and deep learning approaches to combine behavioral analytics with network traffic classification. Three main components make up the architecture: (1) LSTM-based autoencoders for behavioral anomaly identification; (2) a 1D Convolutional Neural Network (1D-CNN) for network intrusion detection; and (3) a multi-level fusion framework that combines the two modalities for reliable decision-making.

4.1 BEHAVIORAL ANOMALY DETECTION USING LSTM BASED AUTOENCODERS

The behavioral analysis component is designed to detect deviations from normal user behavior. This module utilizes the ISOT Web Interactions dataset, which provides detailed records of user interactions within web sessions, including keystroke dynamics, mouse movements, and website interaction patterns. Prior to analysis, the raw behavioral data is preprocessed to generate fixed-length temporal representations. These representations consist of four key numerical features: key hold-time, key delay, mouse movement magnitude, and site activity delta, capturing the temporal evolution and characteristics of user behavior.

To effectively model these temporal dependencies and identify subtle anomalies, three variations of LSTM-based autoencoders were explored:

1. Standard LSTM Autoencoder:* This baseline model processes sequential behavioral features, organized as 200-timestep sequences with 4 features per timestep. The architecture consists of an encoder, a bottleneck layer, and a decoder. The encoder, composed of LSTM layers with progressively decreasing units (e.g., 64 to 32), compresses the input sequence into a lower-dimensional representation. This compressed representation is then passed through a dense bottleneck layer. The decoder, mirroring the encoder's architecture with LSTM layers of increasing units, reconstructs the original input sequence. A TimeDistributed Dense layer generates the final output, aiming to replicate the input sequence as closely as possible. The model is trained to minimize the reconstruction error, quantified as the mean squared error (MSE) between the input and the reconstructed sequence, using only normal behavioral sequences. During inference, anomalous behavior is identified by comparing the reconstruction error of a given sequence to a predefined threshold; errors exceeding this threshold are flagged as anomalies.

2. Deep LSTM Autoencoder:* To capture more intricate and long-range temporal dependencies within user behavior, a deeper LSTM autoencoder architecture was investigated. This enhancement involves incorporating additional stacked LSTM layers in both the encoder and decoder. For instance, the encoder may consist of LSTM layers with units decreasing from 128 to 64 and then to 32, while the decoder mirrors this with increasing units (e.g., 32 to 64 to 128). The inclusion of these additional layers allows the model to learn more abstract and complex representations of behavioral patterns, potentially leading to improved anomaly detection performance.

3. Attention-Based LSTM Autoencoder:* Recognizing that not all time steps within a behavioral sequence contribute equally to the identification of anomalies, an attention mechanism was integrated into the LSTM autoencoder. This enhancement allows the model to selectively focus on the most salient time steps. The encoder comprises LSTM layers followed by attention layers. The attention mechanism generates a context vector, derived from the attention layer outputs, which represents a weighted combination of the LSTM outputs, emphasizing the most relevant temporal information. The decoder also incorporates attention mechanisms. This architecture enables the model to weigh the importance of different time steps during both the encoding and decoding processes. By focusing on critical time steps, the attention-based LSTM autoencoder is better equipped to detect subtle anomalies that might be missed by models that treat all time steps equally.

4.2 NETWORK INTRUSION DETECTION USING 1D CONVOLUTIONAL NEURAL NETWORK

The network intrusion detection component is responsible for analyzing network traffic data to identify malicious activity. This module utilizes the NSL-KDD dataset, a benchmark dataset containing 22,544 labeled network connection samples, each characterized by 41 features. Prior to processing, the network data undergoes preprocessing, including label encoding to convert categorical labels into numerical values, and normalization to scale the features to a common range. The preprocessed data is then reshaped into a format suitable for input into a 1D-CNN, with each sample represented as a 41-dimensional vector with a single channel.

The 1D-CNN architecture is designed to extract relevant patterns from the network traffic features. The architecture consists of the following layers:

Conv1D layer:* A one-dimensional convolutional layer with 32 filters and a kernel size of 3, applied to the input features. This layer is followed by a ReLU (Rectified Linear Unit) activation function to introduce non-linearity.

MaxPooling1D layer:* A max pooling layer to reduce the dimensionality of the feature maps, retaining the most salient information.
Conv1D layer:* Another one-dimensional convolutional layer with 64 filters and a kernel size of 3, again followed by ReLU activation.
MaxPooling1D layer:* Another max pooling layer.
Flatten layer:* A flattening layer to transform the multi-dimensional feature maps into a one-dimensional vector.
Dense layer:* A fully connected (dense) layer with 64 units and ReLU activation.
Dropout layer:* A dropout layer with a dropout rate of 0.5 to prevent overfitting by randomly dropping out a fraction of the neurons during training.
Output layer:* A final dense layer with a single unit and a sigmoid activation function, producing a probability score representing the likelihood of an intrusion.

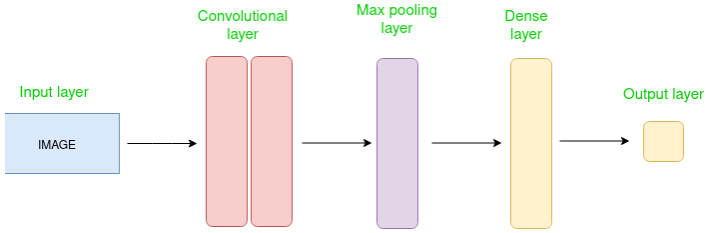

Figure 1: 1D CNN architecture

The 1D-CNN is trained using binary cross-entropy loss, appropriate for binary classification problems, and optimized using the Adam optimizer, an efficient and widely used algorithm for training neural networks. The rationale for using a 1D-CNN lies in its ability to effectively capture local patterns and correlations within the sequential arrangement of network traffic features. This capability makes it well-suited for identifying characteristic signatures of various network intrusion types.

4.3 MULTILEVEL FUSION STRATEGY

To effectively combine the complementary strengths of the behavioral and network-based intrusion detection models, a multi-level fusion strategy is employed. This strategy integrates information from both modalities at multiple stages, encompassing score-level fusion, meta-learning, and ensemble techniques. This approach aims to achieve a more robust and accurate intrusion detection system than could be obtained by relying on either modality alone. The fusion strategy consists of the following components:

1. Meta Neural Network (Meta NN):* This neural network serves as a meta-learner, combining the outputs of the individual behavioral and network models. The input to the Meta NN consists of the concatenated outputs from these models. For example, in the case of the LSTM autoencoders, the input would include the reconstruction errors, representing the degree of behavioral anomaly, and from the 1D-CNN, the probability scores, indicating the likelihood of a network intrusion. The Meta NN architecture comprises several dense layers with ReLU activation functions. Specifically, it includes a dense layer with 8 units, followed by a dropout layer with a rate of 0.2 to mitigate overfitting, and then a dense layer with 4 units, and a final dense layer with 1 unit and sigmoid activation to produce a unified intrusion detection decision. The Meta NN is designed to learn the complex relationships and dependencies between the different modalities, enabling it to make a more informed and accurate assessment of the overall intrusion risk.

2. Super Meta Neural Network:* To further enhance the fusion process and capture higher-order interactions between the modalities, a second-level meta-learner, termed the Super Meta Neural Network, is employed.

The input to this network includes the output of the Meta NN, along with additional relevant features, such as attention scores from the attention-based LSTM autoencoder, if applicable. The architecture of the Super Meta NN is similar to that of the Meta NN, consisting of dense layers with ReLU activation, dropout layers (e.g., with a rate of 0.2), and a final sigmoid output layer. The Super Meta NN aims to model more intricate, non-linear relationships between the fused modalities, potentially leading to improved detection performance, particularly in complex attack scenarios.

3. Weighted Voting:* As a simpler baseline fusion method, a weighted voting mechanism is also implemented. This approach aggregates the binary predictions (intrusion or no intrusion) from the individual behavioral and network models based on predefined weights. For instance, the behavioral model might be assigned a weight of 0.4, while the network model receives a weight of 0.6. The final decision is made by combining the weighted votes. While straightforward, this method lacks the adaptability of the meta-learning approaches, as the weights are fixed and not learned from the data.

4. LightGBM Ensemble Fusion:* The final component of the multi-level fusion strategy is a LightGBM classifier. LightGBM (Light Gradient Boosting Machine) is a gradient boosting framework known for its efficiency and effectiveness. In this context, LightGBM serves as a powerful ensemble learner, combining the outputs of all the preceding stages. The input features to the LightGBM classifier include the outputs from the behavioral and network models (e.g., reconstruction errors and CNN probabilities), the predictions from the Meta NN and Super Meta NN, and the result of the weighted voting. The LightGBM model is trained using 5-fold stratified cross-validation, a robust technique for evaluating model performance and preventing overfitting. Gradient boosting methods like LightGBM are well-suited for handling heterogeneous and non-linear feature combinations, enabling the system to leverage the full information content of the fused data and achieve a high degree of accuracy and robustness.

4.4 Architectural Overview

The overall architecture of the proposed hybrid intrusion detection system can be summarized as follows:

1. Behavioral Pathway:* The input, consisting of raw behavioral data, undergoes preprocessing and is then fed into the LSTM autoencoder variants (Standard, Deep, or Attention-Based). The output of this pathway is a set of anomaly scores, representing the degree of deviation from normal behavior.

2. Network Pathway:* The input, comprising network traffic data, is preprocessed and then analyzed by the 1D-CNN. The output of this pathway is a probability score, indicating the likelihood of a network intrusion.

3. Fusion Pipeline:* The outputs from the behavioral and network pathways (anomaly scores and probability scores) are combined and processed through the multi-level fusion framework. This involves the Meta NN, the Super Meta NN, and the LightGBM classifier, culminating in a final, integrated intrusion detection decision.

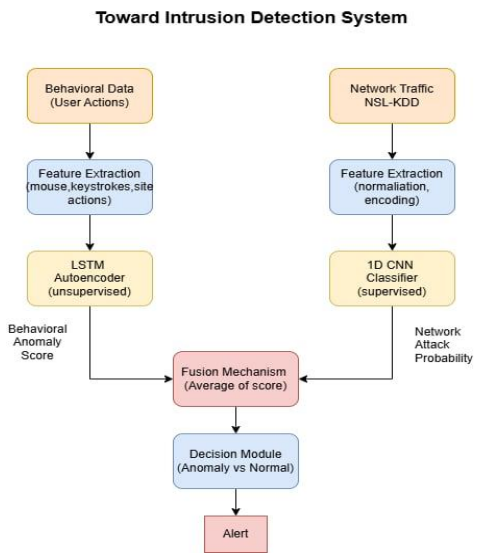

Figure 2: System Architecture

5. RESULT

This section presents the evaluation of the proposed multi-modal intrusion detection framework, which integrates network-level and behavioral-level indicators to enhance threat detection. The evaluation is divided into three parts: performance of behavioral models using the ISOT dataset, performance of the network model using the NSL-KDD dataset, and comparative performance of multiple fusion strategies that combine both modalities.

5.1. Behavioral-Based Intrusion Detection

To detect anomalous user behavior, three distinct LSTM-based autoencoder architectures were trained using only genuine user sessions from the ISOT Web Interactions dataset. Each model was evaluated on 99 attack sessions using a threshold-based anomaly scoring approach derived from reconstruction error.

1. **Standard LSTM Autoencoder**
   A basic encoder–decoder model with two LSTM layers (64 → 32 units), trained to reconstruct the original sequence. Anomalies are identified when reconstruction error exceeds a percentile-based threshold. Performance**:** Precision = 1.0000, Recall = 0.8990, F1-Score = 0.9468

2. **Deep LSTM Autoencoder**
   A deeper variant incorporating three stacked LSTM layers (128 → 64 → 32) to capture richer temporal dependencies in user behavior. Performance**:** Precision = 1.0000, Recall = 0.8990, F1-Score = 0.9468

3. **Attention-Based LSTM Autoencoder**
   Augments the LSTM encoder with a self-attention mechanism to weigh temporal steps based on relevance. The decoder mirrors the encoder and reconstructs the input sequence.
   Performance: Precision = 1.0000, Recall = 0.8990, F1-Score = 0.9468

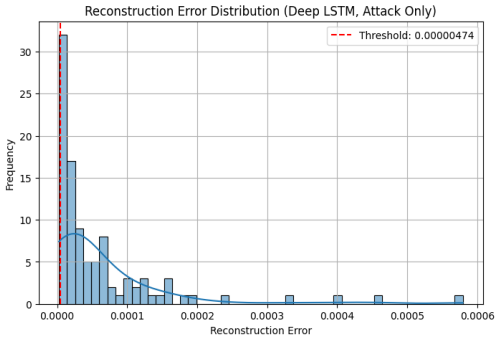

Figure 3: Reconstruction Error

Classification Report:

	precision	recall	f1-score	support
0	0.0000	0.0000	0.0000	0
1	1.0000	0.8990	0.9468	99
accuracy			0.8990	99
macro avg	0.5000	0.4495	0.4734	99
weighted avg	1.0000	0.8990	0.9468	99

Figure 4: Classification Report (Deep LSTM)

Observation: All three behavioral models demonstrated identical performance with extremely high precision and robust recall. However, due to the limited number of labeled attack sessions, their standalone generalizability remains constrained.

5.2. Network-Based Intrusion Detection

The network-level model was trained using the NSL-KDD dataset consisting of 22,544 labeled samples. A 1D Convolutional Neural Network (CNN) was designed to learn spatial patterns from normalized and encoded network traffic features.

4. **1D CNN Model**

   The architecture comprises two Conv1D layers (32 and 64 filters respectively), interleaved with max-pooling, followed by a dense and dropout layer before the final sigmoid output. The model was optimized with binary cross-entropy and Adam optimizer. **Performance:** Precision = 0.9681, Recall = 0.6118, F1-Score = 0.7497

Classification Report:

	precision	recall	f1-score	support
0	0.6548	0.9733	0.7829	9711
1	0.9681	0.6118	0.7497	12833
accuracy			0.7675	22544
macro avg	0.8115	0.7926	0.7663	22544
weighted avg	0.8331	0.7675	0.7640	22544

Figure 5: Classification Report (1D CNN)

Observation: The CNN exhibited high precision, confirming its strength in identifying attacks. However, the relatively low recall points to a high false negative rate.

5.3. Fusion-Based Intrusion Detection

To leverage the strengths of both behavioral and network-based detection, multiple fusion strategies were explored:

1. **Meta Neural Network Fusion**

   A shallow feedforward neural network was trained using normalized reconstruction errors from the ISOT autoencoders and prediction probabilities from the CNN. This approach achieved a precision of 0.9462, recall of 0.8013, and an F1-score of 0.8677. The significant improvement in recall compared to the CNN-only model highlights the benefit of incorporating behavioral signals in reducing false negatives.

2. **Super Meta Neural Network**

An enhanced version of the meta neural network was developed by stacking an additional neural layer that combined predictions from the attention-based autoencoder and the meta neural network. This model achieved a slightly higher recall (0.8236) while maintaining strong precision (0.9434), resulting in an improved F1-score (0.8794). The super meta neural network demonstrated the most balanced trade-off between false positives and false negatives among the fusion methods tested.

3. Weighted Voting

A simpler fusion strategy was implemented by aggregating binary predictions from the ISOT and NSL-KDD models using weighted voting (0.4 for behavioral, 0.6 for network). However, this approach did not yield any performance improvement, matching the F1-score (0.7497) of the standalone CNN model.

4. LightGBM Fusion Model

To further optimize detection performance, a LightGBM classifier was trained on a five-dimensional feature vector comprising attention-based reconstruction scores, meta neural network outputs, weighted voting results, and raw prediction scores from both ISOT and NSL-KDD models. Using five-fold stratified cross-validation, the LightGBM fusion model achieved the best overall performance, with precision (0.9014), recall (0.9050), and F1-score (0.9032), along with an accuracy of 88.96%. This model effectively balanced precision and recall, making it the most robust fusion strategy.

```
📊 Final LightGBM Fusion Evaluation (on all 22544 samples):
Precision: 0.9014
Recall:    0.9050
F1 Score:  0.9032

Confusion Matrix:
[[ 8441  1270]
 [ 1219 11614]]

Classification Report:
              precision    recall  f1-score   support

           0     0.8738    0.8692    0.8715      9711
           1     0.9014    0.9050    0.9032     12833

    accuracy                         0.8896     22544
   macro avg     0.8876    0.8871    0.8874     22544
weighted avg     0.8895    0.8896    0.8896     22544
```

Figure 6: Classification Report (LightGBM)

5.4. Feature Importance Analysis

Feature importance analysis was conducted on the final LightGBM fusion model using gain-based evaluation metrics. The results indicated that the most influential feature contributing to accurate intrusion detection was the raw probability output generated by the 1D Convolutional Neural Network trained on the NSL-KDD dataset. This underscores the central role of network-level data in identifying malicious activity. The second most significant feature was the output score from the meta neural network, which captures interactions between behavioral and network indicators, suggesting that learned fusion enhances model expressiveness. In contrast, behavioral indicators derived from the ISOT dataset—such as the reconstruction errors from the attention-based and standard LSTM autoencoders—had relatively lower importance. This discrepancy is likely attributable to the limited number of attack samples in the behavioral dataset, which constrained the learning signal. Nonetheless, the inclusion of these features still contributed to the model's robustness, confirming the value of multi-modal inputs in complex intrusion detection scenarios.

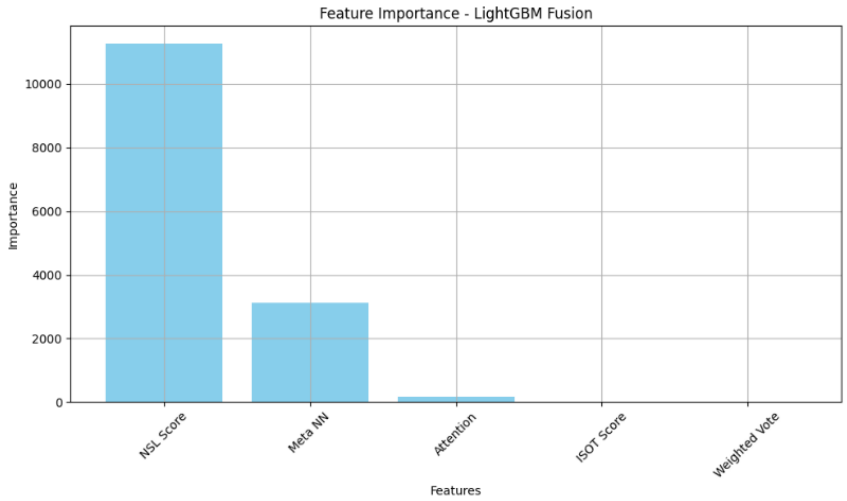

Figure 7: Feature importance

5.5. Summary of Analysis

The comparative performance of all evaluated models is summarized below:

Model	Precision	Recall	F1-Score
CNN (NSL-KDD)	0.9681	0.6118	0.7497
Behavioral (ISOT Autoencoder)	1.0000	0.8990	0.9468
Meta Neural Network	0.9462	0.8013	0.8677
Super Meta Neural Network	0.9434	0.8236	0.8794
LightGBM Fusion	0.9014	0.9050	0.9032

Table 1: Model Score Comparison

The results indicate that while the network-based CNN provides strong precision, its standalone recall is insufficient for comprehensive threat detection. Behavioral models, despite their high precision and recall on limited data, may not generalize well without additional training samples. Fusion strategies, particularly the LightGBM-based approach, successfully combine the strengths of both modalities, achieving the most balanced and scalable intrusion detection performance.

6. DISCUSSION

The experimental results demonstrate several key insights into the effectiveness of the proposed multi-modal intrusion detection framework:

1. **Behavioral Models Are Highly Precise but Limited by Dataset Size**

   All three LSTM-based behavioral models (standard, deep, and attention) achieved perfect precision (1.0000) and a recall of 0.8990. This indicates a strong ability to avoid false positives while accurately detecting a high proportion of attacks within the ISOT dataset. However, the evaluation was constrained to only 99 attack sessions, highlighting a limitation in generalizability due to the small behavioral test set.

2. **Network Model Provides Strong Baseline but Suffers from Low Recall**

   The 1D CNN trained on the NSL-KDD dataset achieved high precision (0.9681) but comparatively lower recall (0.6118), revealing its tendency to misclassify many attacks as benign. This underlines the vulnerability of relying solely on network data, especially against sophisticated or less frequent attack patterns.

3. **Fusion Improves Recall and Overall Detection Performance**

   The Meta Neural Network significantly improved recall (0.8013) over the standalone CNN while maintaining high precision. This indicates that incorporating behavioral signals—despite being limited in quantity—adds valuable context for distinguishing attacks.

4. **Super Meta Neural Network Provides Better Balance**

   The stacked model combining attention-based outputs and Meta NN predictions achieved a balanced performance (Precision: 0.9434, Recall: 0.8236, F1: 0.8794), showing that higher-order feature interactions across modalities enhance robustness.

5. **LightGBM Fusion Achieves the Best Overall Performance**

   The LightGBM model, trained on five fusion-derived features, delivered the highest F1 score (0.9032) and balanced precision/recall, proving that gradient boosting can effectively model complex relationships in multi-modal data.

6. **Behavioral Features Contribute, But Network Features Dominate**

   Feature importance analysis revealed that while behavioral outputs contributed meaningfully, the raw CNN predictions and meta-fusion scores had the most impact. This suggests that behavioral data complements—but does not replace—the more comprehensive coverage of network traffic data.

7. **Weighted Voting Was Ineffective Compared to Learning-Based Fusion**

   The simple weighted voting method mirrored the baseline CNN performance, reaffirming that static fusion strategies lack the adaptability required for nuanced threat detection.

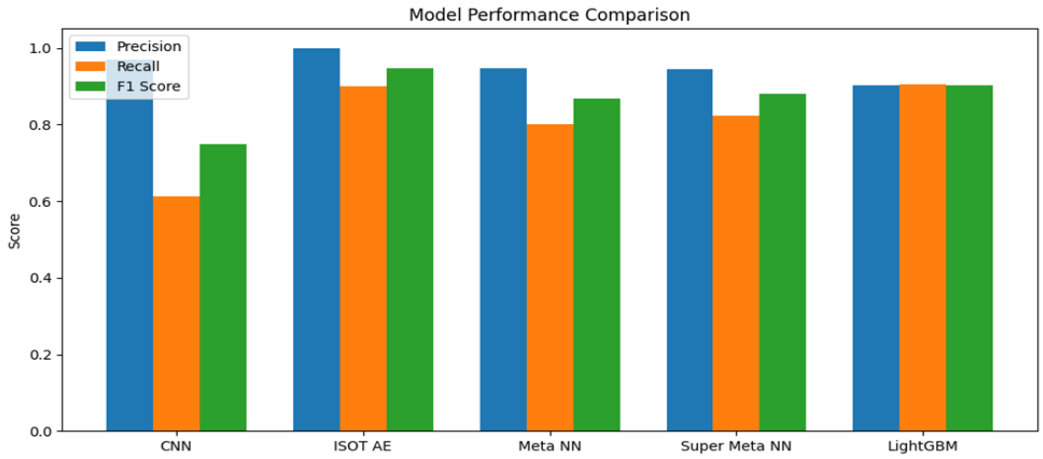

Figure 8: Model Performance

7. FUTURE SCOPE

This research establishes the foundation for multi-modal intrusion detection, but several enhancements are planned:

1) **Dataset Expansion**: The behavioral model performance is currently constrained by the limited size and imbalance of the ISOT dataset. Future work will incorporate larger, more diverse datasets with richer user interaction logs and varied attack patterns to improve generalization.

2) **Real-Time Detection**: Adapting the framework for online, real-time monitoring will be a key objective, enabling timely responses to evolving threats in operational environments.

3) **Advanced Fusion Strategies**: Future research will explore adaptive or attention-based fusion mechanisms to dynamically weigh the importance of behavioral and network inputs based on context.

4) **Explainability and Robustness**: Integrating explainable AI techniques and improving resistance to adversarial inputs will make the system more transparent and resilient for practical deployment.

5) **Cross-Session Behavioral Modeling**: Incorporating sequential learning across multiple sessions could enhance the detection of low-frequency, stealthy intrusions.

8. CONCLUSION

This study presented a comprehensive multi-modal intrusion detection framework that integrates behavioral and network-based analytics to enhance threat detection accuracy. By combining LSTM-based autoencoders trained on user interaction data (ISOT Web Interactions) with a 1D-CNN trained on network traffic (NSL-KDD), and further refining decisions through multiple fusion strategies—including meta learning and LightGBM ensembles—the system demonstrates a significant improvement over single-modality approaches. Experimental results confirm that fusion not only mitigates the weaknesses of individual models but also yields better precision-recall trade-offs. Notably, the LightGBM fusion achieved the highest overall F1-score of 0.9032, proving the effectiveness of ensemble learning in multi-source intrusion detection. This work underscores the value of integrating heterogeneous data sources and learning-based fusion techniques to build more robust, adaptive, and intelligent cybersecurity systems.

REFERENCES

[1] Ahmed, I., Traore, I., & Ahmed, A. (2017). A survey of behavioural biometrics and their challenges. Information Fusion, 34, 63-71.

[2] Bleha, J., Obdrzalek, D., & Mach, M. (2017). Keystroke dynamics authentication: A survey. Computers & Security, 68, 191-212.

[3] Hindy, H. A., Brosset, D., & Boudy, J. (2018). A survey on network intrusion detection in IoT. IEEE Communications Surveys & Tutorials, 20(3), 1646-1669.

[4] Kim, J., Kim, H., & Kim, K. (2020). Deep learning-based network intrusion detection: A survey. International Journal of Advanced Computer Science and Applications, 11(1), 39-55. [5] Tang, T. A., McLernon, D., Zaidi, S. A. R., & Ghogho, M. H. (2017). CNN-based intrusion detection system for wireless networks. IEEE Access, 5, 3515-3525.

[6] Vinayakumar, R., Alazab, M., Soman, K. P., Poornachandra, S., Al-Nemrat, A., & Azeez, A. (2019). Applying deep learning approaches for network intrusion detection. IEEE Access, 7, 4727-4747.

[7] Khan, M. A., Liyanage, M., Okwuibe, J., & Ylianttila, M. (2021). Federated learning for intrusion detection in industrial internet of things. IEEE Access, 9, 11772-11782.

[8] Liu, X., Huang, C., & Zhang, D. (2018). Graph-based anomaly detection for identifying malicious activities in enterprise networks. In Proceedings of the 2018 World Wide Web Conference (pp. 1041-1050).

[9] Sharafaldin, I., Lashkari, A. H., Saeedi, R., & Ghorbani, A. A. (2021). Explainable artificial intelligence for cyber security: Opportunities, challenges, and research directions. In 2021 International Carnahan Conference on Security Technology (ICCST) (pp. 1-8). IEEE.

[10] Agrawal, M., & Thakkar, A. (2020). Cloud Intrusion Detection System using Machine Learning Techniques. International Journal of Innovative Technology and Exploring Engineering (IJITEE), 9(4), 2278-3075.

[11] Bostani, H., Sheykhian, A., & Razavi, S. N. (2017). Intrusion detection system based on deep neural network and fuzzy logic. Journal of Intelligent & Fuzzy Systems, 33(6), 3939-3950.

[12] Javaid, N., Niyaz, Q., Sun, W., & Alam, M. (2016). A deep learning framework for cyber threat intelligence knowledge extraction. In 2016 IEEE International Conference on Cyber Security and Cloud Computing (CSCloud) (pp. 129-134). IEEE.

[13] Lopez-Martin, M., Garcia-Domínguez, J., & Lopez, J. M. (2017). Network traffic classification using deep learning. In 2017 Seventh International Conference on the Innovative Internet Community Services (I2CS) (pp. 129-136). IEEE.

[14] Marchal, S., Armknecht, F., State, R., & Fischer, M. (2017). A survey of attacks on network infrastructure. IEEE Communications Surveys & Tutorials, 19(1), 730-757.

[15] Moustafa, N., & Slay, J. (2017). UNSW-NB15: A comprehensive data set for network intrusion detection systems (NSW-NB15 network data set). In 2015 Military Communications and Information Systems Conference (MilCIS) (pp. 1-6). IEEE.

[16] Shafi, M., Liu, X., & Yu, P. S. (2020). Targeted Adversarial Attacks on Deep Learning-Based Network Intrusion Detection Systems. IEEE Transactions on Dependable and Secure Computing, 17(6), 1161-1174.

[17] Thamilarasan, P., & Kuppusamy, E. (2019). Deep Neural Network based Intrusion Detection System for Internet of Things. International Journal of Innovative Technology and Exploring Engineering (IJITEE), 8(12), 2788-2792.

[18] Wang, W., Sheng, Y., Wang, J., Chen, J., & Zeng, X. (2018). A survey of machine learning for network anomaly detection. IEEE Access, 6, 3525-3547.

[19] Xia, Y., Qu, G., Li, J., & Gu, X. (2020). Evasion Attacks against Deep Learning based Network Intrusion Detection Systems. In Proceedings of the 13th ACM Conference on Security and Privacy in Wireless and Mobile Networks (pp. 209-219).

[20] Zuech, R., Khawaja, B. A., Abdullah, H., & Javaid, N. (2020). A Survey on the Applications of Deep Learning in Intrusion Detection Systems. IEEE Access, 8, 181138-181154.

[21] Almiani, M., Abu-Ali, N., El-Khatib, K., & Sharieh, A. (2020). A systematic survey of deep learning-based network intrusion detection systems. Computers & Security, 97, 101990.

[22] Buczak, A. L., & Guven, E. (2016). A survey of data mining and machine learning methods for cyber security intrusion detection. IEEE Communications Surveys & Tutorials, 18(2), 1037-1070.

[23] Garcia, N., & Plattner, B. (2020). Anomaly-based network intrusion detection with self-organizing maps. In Artificial Neural Networks and Machine Learning–ICANN 2020: Proceedings of the 29th International Conference on Artificial Neural Networks (pp. 1-13). Springer Nature.

[24] Hoonselaar, A., & Šnajder, J. (2021). Evaluating deep learning for network intrusion detection. Journal of Computer Virology and Hacking Techniques, 17(1), 69-92.

[25] Ring, M., Wunderlich, S., Scheuermann, H. K., & Landes, S. (2019). A survey of network-based intrusion detection data sets. Computers & Security, 86, 147-167.