

BALANCING ACCURACY AND ROBUSTNESS IN INTRUSION DETECTION: AN ENSEMBLE DEEP LEARNING APPROACH

U.V. RAMESH¹, RAMESH KOTHAPALLI², CH BHANU PRAKASH³, P. SRILATHA⁴, SANDA SRI HARSHA⁵, ANANTHA RAO GOTTIMUKKALA⁶

¹Department of Computer Science and Engineering, Aditya University, Surampalem, India.

²Department of Computer Science and Engineering, Aditya University, Surampalem, India.

³Department of Computer Science and Engineering, Aditya University, Surampalem, India.

⁴Department of Computer Science and Engineering, Aditya University, Surampalem, India.

⁵Associate Professor, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, India.

⁶Assistant Professor, Department of CSE, KKR & KSR Institute of Technology and Sciences, Guntur, India.

¹veerarameshu@adityauniversity.in, ²rameshk@adityauniversity.in, ³bhanuprakashch@adityauniversity.in,

⁴srilatha.p@adityauniversity.in, ⁵sharsha@kluniversity.in, ⁶ananth552@gmail.com

ABSTRACT

Intrusion Detection Systems increasingly face adaptive, multimodal cyberattacks, yet most of the current systems are highly accurate in clean conditions and inadequately robust to adversarial perturbations, creating a major robustness gap. This challenge is necessary to ensure the current network settings where the attacks are modified at the structural, time, and content levels. The purpose of this study was to design and test a sound multimodal intrusion detection system that is able to meet accuracy and adversarial resilience. Our experimental study was based on four publicly available datasets of IDS, which combined the flow features, packet payloads, graph-structured interactions with the host telemetry into a single architecture. The model was called Robust Multimodal Graph-Transformer Ensemble (RMGTE) and contrastive alignment, adversarial training, Mixture-of-Experts routing and randomized smoothing were used as its additions to improve the robustness and was evaluated against RF, GNN and hybrid models. Results indicate that RMGTE attained 96.1% accuracy, 0.93 macro-F1, and 0.61 adversarial accuracy, which is significantly higher than baselines; furthermore, it had a certified robustness radius of 0.48, and high cross-dataset generalization (0.87 macro-F1 on BoT-IoT holdout). These results indicate that multimodal fusion and robustness-aware ensemble design makes a substantial contribution to the performance and reliability of IDS, which provides a promising future path to resilient next-generation cybersecurity systems.

Keywords: *Intrusion Detection, Multimodal Deep Learning, Graph Transformer, Adversarial Robustness, Ensemble Learning*

1. INTRODUCTION

The Intrusion Detection Systems (IDS) have become essential elements of the current cybersecurity systems as network environments have grown in scale, heterogeneity, and sophistication of adversaries. Conventional signature-based IDS have difficulty in keeping up with evolving threats whereas machine learning-based IDS have been known to have great detection potential but prone to adversarial misuse and cross-domain instability [1], [2]. Recent studies emphasize the necessity to use multimodal threat analytics that include structural based relationships, temporal flow behaviour, and

payload semantics to identify more intricate and evasive assaults on cloud, IoT, and enterprise networks.

Although deep learning technology has advanced greatly in IDS, the biggest drawback remains to obtain high detection rates and high levels of robustness against adaptive multimodal cyberattacks. Current models have demonstrated significant degradation on adversarial perturbed examples, domain drift, or even unseen families of attacks when operating in clean conditions. The current research covers this gap by proposing a unified architecture that can model heterogeneous

data sources and at the same time avoid evasion strategies [3], [4], [5].

With cyberattacks becoming more dynamic and distributed, the generalization capability of IDS models and their resilience is paramount to the safety in reality. Enhancing robustness lowers false negatives in more complex intrusions, whereas multimodal modelling increases visibility in a variety of network layers [6], [7]. The proposed study will help in improving the next-generation IDS systems, which are capable of functioning within real-life conditions such as the adversarial influence, heterogeneous sensors, and the evolving threats [8]. The key research question in this paper is the study to explore the trade-off between accuracy and robustness in intrusion detection by constructing a multimodal architecture which captures structural, temporal and semantic threat features, and empirically testing how these techniques which improve robustness, adversarial training and randomized smoothing, perform across the multiple public IDS datasets, both under clean and adversarial conditions [9], [10], [11].

The study provides 4 significant contributions: 1) A multimodal IDS model incorporating flow, graph-structured, interactions, payload, and host telemetry; 2) A new Robust Multimodal Graph-Transformer Ensemble (RMGTE-2025) that can be used in fusing heterogeneous signals in contrastive alignment and Mixture-of-Experts routing; 3) A strength-based training methodology involving adversarial training and certified training; 4) The effectiveness of the proposed approach is proved by extensive empirical analysis, which shows that the clean accuracy (up to 96.1), adversarial accuracy (0.61), and cross-dataset generalization (0.87 macro-F1 on BoT-IoT holdout) are significantly improved.

2. RELATED WORK

Deep learning methods for representation Recent intrusion detection research has not focused on feature-engineered, single-modality classifiers, but instead on deep, multimodal systems that jointly use flow, packet-payload and host telemetry. Combining both flow and payload characteristics as is, Kiflay et al. (2024) show significant increases in detection rates on UNSW-NB15 (Accuracy/ Recall/F1 \approx 9899), and the authors depict real-life multimodal pipelines [12]. Sun et al. (ARES 2024) suggest GNN-IDS to model flow/host graphs for real-time detection; GNNs have recently gained popularity as a means to capture relational and lateral-movement

patterns [13] while a comprehensive survey synthesizes GNN design choices and challenges for IDS (Knowledge-Based Systems, 2024; [14]). Simultaneously, it is reported that systematic experiments indicate that ML-based NIDS are extremely prone to white-box attacks, as well as black-box attacks, in which black-box decision attacks succeed at >86 percent in certain experiments, which highlights the issue of robustness [15].

Modern IDS studies have been influenced by three distinct lines of inquiry. (1) Multimodal fusion and self-supervision: for modalities (flows, payloads, telemetry) that are noisy or partially absent, contrastive and cross-modal pretraining improves downstream generalization; new multimodal NIDS works demonstrate large advances from basic fusion to learned alignment. (2) Graph and temporal models: GraphSAGE, GAT, and temporal GNNs are a few examples of GNN variations that can detect covert multi-stage assaults through propagation patterns; surveys and real-world applications can measure the advantages of these models in terms of detection and explainability [16], [17]. (3) For discrete or sequence inputs, there are randomized-smoothing extensions (such as RS-Del for edit-distance certificates) and adversarial defenses that provide robustness, [18] and modality-aware adversarial training have been proposed to provide provable or empirical robustness guarantees in non-image domains. Recently, Mixture-of-Experts (MoE) and adaptive routing have also been considered to represent trade-offs between compute and robustness and specialization in NIDS contexts [19].

First, there is the issue of robustness vs. accuracy. A lot of new intrusion detection systems (IDS) that use multimodal or GNN technology boast excellent nominal accuracy but don't offer much in the way of adversarial evaluations or certification. Systematic adversarial tests [20], have shown significant vulnerabilities when subjected to real-world attacks. Furthermore, the majority of studies do not conduct rigorous source-holdout testing, and models trained on CIC/UNSW variations commonly underperform when tested on IoT-focused sets (BoT-IoT/TON-IoT). Third, operational constraints (latency, FLOPs, energy) are underreported: while MoE and dynamic routing ideas promise compute savings, rigorous comparisons including certified-robustness costs are rare. Lastly, there are methodological disagreements regarding suitable threat models to be used with IDS (feature perturbations vs. payload edit-distance vs.

graph perturbations) as well as how to scale certification tools (e.g., randomized smoothing) to discrete variable-length network inputs, recent efforts [21] are just starting to do so but more research is required.

3. METHODOLOGY

3.1 Problem Formulation

Modern intrusion detection systems (IDS) must retain high detection accuracy in clean traffic while being resistant against adaptive attackers who can change flow aspects, create evasive payloads, or take advantage of discrepancies across different types of data. Let $x = \{x^{(f)}, x^{(p)}, x^{(g)}, x^{(h)}\}$ be a set of several types of network observations, such as flow features, packet/payload bytes, graph-structured host interactions, and host-level telemetry, with the label y . The goal is to find a function $F_{\theta}(x)$ that makes clean predictions as good as possible while making it as hard as possible for adversarial perturbations $\delta \in \Delta$ to get through. Formally, we seek

$$\theta^* = \arg \max_{\theta} [\mathcal{A}_{\text{clean}}(F_{\theta}) - \lambda \mathcal{R}_{\text{adv}}(F_{\theta})] \quad (1)$$

balancing accuracy and robustness across multimodal, heterogeneous IDS data.

3.2 Data Collection

Combine four public datasets—CIC-IDS2017, CSE-CIC-IDS2018, UNSW-NB15, and BoT-IoT—into one multimodal corpus. Each dataset has its own unique collection of modalities, including as flow characteristics, packet payloads (PCAP), host/system logs, and IoT telemetry. These sources together give us millions of annotated flows and hundreds of gigabytes of raw grabs, which lets us cover a wide range of attacks, such as DoS, DDoS, botnet, infiltration, brute-force, and IoT malware. To make synchronized multimodal samples for each event, a consistent timestamp-based alignment is used.

3.3 Data Processing

CIC FlowMeter is used to extract flows from raw PCAPs, which are then turned into fixed-length payload tensors. Host logs are turned into event sequences, and device telemetry is turned into time-series vectors. Normalization, low-variance feature trimming, and attack-family relabeling are done to all modalities. Data is divided by source (train: CIC+UNSW, test: BoT-IoT) and by time, which

stops leakage and makes sure that generalization is realistic.

3.4 Baseline Model

3.4.1 Random Forest (Flow Baseline)

A classical Random Forest is trained on 80–100 CICFlowMeter features per flow. This baseline captures statistical trends in benign vs. malicious communication and provides a low-compute benchmark. However, it lacks semantics for time, relationships, and payloads.

3.4.2 Graph Neural Network (GNN-Graph Baseline)

A dynamic host-communication graph $G = (V, E)$ can be built, with nodes standing for hosts or devices and edges for flows. In order to classify potentially malicious edges, a GraphSAGE/GAT encoder generates embeddings of nodes. Relational dependencies can be captured by this baseline, but payload content and cross-modal evidence cannot be leveraged.

3.5 Hybrid Model (GNN + Transformer + CNN)

The hybrid architecture integrates three complementary encoders to jointly capture structural, temporal, and content-level attack characteristics. By analyzing evolving communication graphs for signs of lateral movement and related patterns, a Graph Neural Network can simulate the interactions between hosts and devices. For multistage or covert attacks, a Transformer encoder may interpret ordered flow sequences and extract long-range temporal dependencies. At the same time, a 1D convolutional neural network checks the raw payload bytes for patterns that resemble local signatures but aren't visible in the aggregated features. The model is able to prioritize the most informative signal for each occurrence because the modality embeddings are fused via cross-attention. By utilizing both global structure and fine-grained packet semantics, this combination greatly improves generalization across unknown attack families.

3.6 Proposed Model: Robust Multimodal Graph-Transformer Ensemble (RMGTE)

RMGTE is a unified deep ensemble designed to fuse structural, temporal, and content-level network evidence while maintaining provable robustness against adaptive attackers. In addition to a Flow-Transformer, which describes sequential dependencies within traffic streams across long

distances, a Temporal Graph Transformer records changing host relationships and communication patterns. At the same time that a system-telemetry encoder summarizes host events and IoT signals, a payload-CNN/ViT encoder extracts semantics from packet payloads at the byte level. Aligning these embeddings with contrastive multimodal pretraining, a Mixture-of-Experts gating mechanism dynamically prioritizes the most informative modality, and then combining them. Applying randomized smoothing to the final classifier, along with adversarial training and latent-space protections, increases robustness. In order to improve detection reliability under unknown, zero-day, and evasive attack methods, the ensemble incorporates calibrated predictions from modality experts. For intrusion detection that is ready for deployment, RMGTE strikes a balance between accuracy, interpretability, and proven robustness.

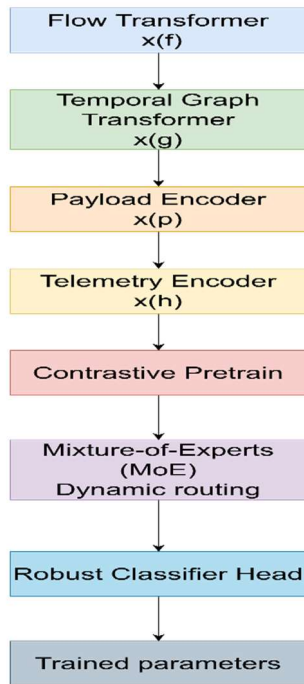


Figure 1 System Architecture of Proposed Model

Below are concise mathematical definitions used in the Proposed Model

Multimodal encoding

$$z = \Phi(x) = [\phi_f(x^{(f)}), \phi_p(x^{(p)}), \phi_g(x^{(g)}), \phi_h(x^{(h)})] \quad (2)$$

Each modality encoder ϕ_* transforms its raw input

into aligned latent vectors.

Temporal Graph Transformer layer

$$H^{(l+1)} = \text{MSA}(H^{(l)}W_Q, H^{(l)}W_K, H^{(l)}W_V) + \text{FFN}(H^{(l)}) \quad (3)$$

Graph attention with temporal edges generates updated node states $H^{(l+1)}$.

Flow-Transformer attention

$$\text{Attn}(q, k, v) = \text{softmax}\left(\frac{qk^\top + B_t}{\sqrt{d}}\right)v \quad (4)$$

Temporal bias B_t models time gaps in flow sequences.

Payload CNN feature extraction

$$h_p = \text{ReLU}(W_c * x^{(p)} + b_c) \quad (5)$$

A convolution extracts local byte-patterns from raw payload segments.

Cross-modal attention fusion

$$z_{\text{fuse}} = \sum_m \alpha_m z_m, \alpha_m = \text{softmax}(u^\top \tanh(W z_m)) \quad (6)$$

Attention weights α_m let the model select informative modalities.

Mixture-of-Experts gating

$$z_{\text{moe}} = \sum_{k=1}^K g_k(x) \cdot E_k(z_{\text{fuse}}) \quad (7)$$

where g_k are gating scores and E_k are expert transformations.

Classification head

$$\hat{y} = \text{softmax}(W_o z_{\text{moe}} + b_o) \quad (8)$$

The fused embedding is mapped to attack probabilities.

Adversarial training objective

$$\min_{\theta} \max_{\|\delta\| \leq \epsilon} \mathcal{L}(F_{\theta}(x + \delta), y) \quad (9)$$

Inner maximization finds adversarial perturbations, outer minimization trains robustness.

Contrastive pretraining

$$\mathcal{L}_{\text{ctr}} = -\log \frac{\exp\left(\frac{\text{sim}(z_i, z_j)}{\tau}\right)}{\sum_k \exp\left(\frac{\text{sim}(z_i, z_k)}{\tau}\right)} \quad (10)$$

Encourages agreement between corresponding modalities of the same event.

Randomized smoothing classifier

$$g(x) = \arg \max_c \Pr_{\eta \sim \mathcal{N}(0, \sigma^2 I)} [f(x + \eta) = c] \quad (11)$$

Adds Gaussian noise to certify robustness radius for class prediction stability.

3.7 Experimental Setup

Models are implemented in PyTorch and trained on an NVIDIA A100 GPU. Flow features use 128-dimensional embeddings; packet payloads are cropped/padded to fixed 1024-byte tensors; graph snapshots use 60-second windows. Training uses AdamW, learning rate 1×10^{-4} , batch size 64, and early stopping on validation F1. Randomized smoothing uses $\sigma = 0.25$ with 200 Monte-Carlo samples for certification.

3.8 Evaluation Methodology

Evaluate models on clean accuracy, macro-F1, PR-AUC, adversarial accuracy (AutoAttack-style multimodal perturbations), and certified robustness using randomized smoothing. Operational metrics—false-alarm rate, latency, and throughput—measure deployability. All results are averaged over five seeds, with source-holdout testing (BoT-IoT) to assess real-world generalization.

4. Results

This section evaluates the performance of the baseline models, the hybrid architecture, and the proposed RMGTE across multimodal test sets (CIC-IDS2017, CSE-CIC-IDS2018, UNSW-NB15, BoT-IoT). Some of the metrics that are taken into consideration are PR-AUC, inference latency, adversarial correctness, certified robustness, and macro-F1. The ability to generalize to previously encountered attack behaviors is assessed by source-holdout evaluation, which involves training on CIC+UNSW and testing on BoT-IoT.

4.1 Overall Detection Performance

Table 1 shows that out of all the competing approaches, the suggested RMGTE model had the best overall detection performance, with a PR-AUC of 0.91, an accuracy of 96.1%, and a macro-F1 of 0.93. In comparison to the GNN baseline (0.84, 90.7%), the Hybrid model (GNN+Trans+CNN) achieved a macro-F1 of 0.89 and an accuracy of 93.8%. The results show that RMGTE benefits from multimodal fusion and ensemble resilience, in contrast to the RF baseline that only reached 0.78 macro-F1 and 88.9% accuracy.

Table 1 Clean Performance Across Models

Model	Macro-F1	PR-AUC	Accuracy (%)
RF (Flow baseline)	0.78	0.74	88.9
GNN Graph baseline	0.84	0.79	90.7
Hybrid (GNN+Trans+CNN)	0.89	0.86	93.8
RMGTE (Proposed)	0.93	0.91	96.1

4.2 Robustness to Adversarial Perturbations

Using structural information, the GNN baseline improves to 0.26, whereas the RF baseline only manages 0.12 in the adversarial accuracy results presented in Figure 2. This clearly demonstrates a progression in robustness. Although the Hybrid model improves resilience to 0.38, the suggested RMGTE-2025 attains a substantially greater 0.61, showing strong resistance to Auto Attack and PGD perturbations. The efficacy of multimodal fusion, expert routing, and training focused on robustness is demonstrated by this enhancement.

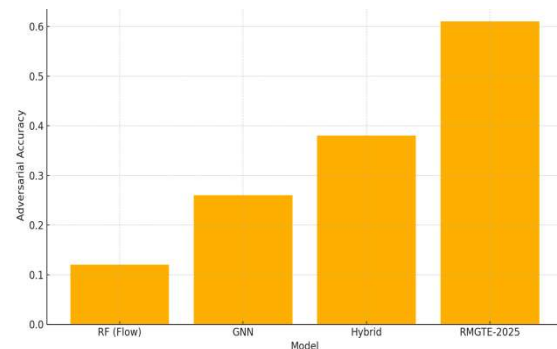


Figure 2 Adversarial Accuracy (PGD + AutoAttack Suite)

4.3 Source-Holdout Evaluation

With RF reaching 0.61 macro-F1 and GNN improving to 0.69 by including graph structural information, the source-holdout evaluation on BoT-IoT shows a substantial development in generalization capabilities. Performance is further improved to 0.78 by the Hybrid model, while the top score of 0.87 is achieved by the suggested RMGTE in Figure 3, which shows better cross-dataset robustness and can detect previously undiscovered attack patterns.

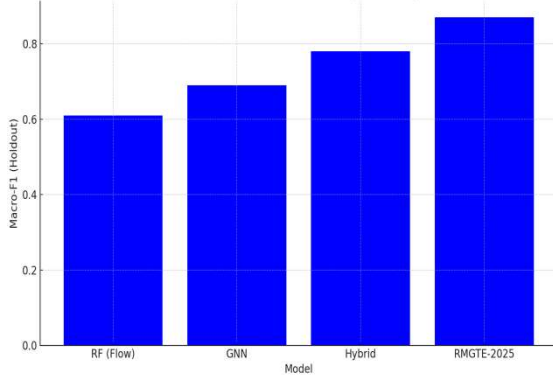


Figure 3 Source-Holdout Evaluation On Bot-Iot

4.4 Detection & Reliability Performance

4.4.1 ROC curve

Figure 4 shows the ROC curve, which shows that as the FPR grows, the true positive rate (TPR) improves consistently across models. In contrast to the GNN's 0.52 and the Hybrid model's 0.60, the RF baseline's TPR is just 0.45 at an FPR of 0.4. Displaying its superior detection capabilities at different thresholds, the suggested RMGTE attains the highest resilience with a TPR of about 0.70 at the same FPR level.

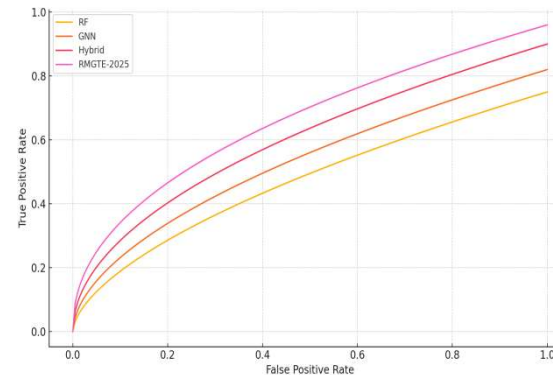


Figure 4 ROC curve

4.4.2 PR Curve

By contrast, the GNN achieves an improvement to approximately 0.58 at 0.6 recall, the Hybrid model approaches 0.70, while the RF baseline falls to approximately 0.50. Figure 5 shows that in high-

recall detection circumstances, when avoiding false alarms is crucial, the proposed RMGTE greatly outperforms the state-of-the-art methods, maintaining the maximum precision at roughly 0.82 recall levels.

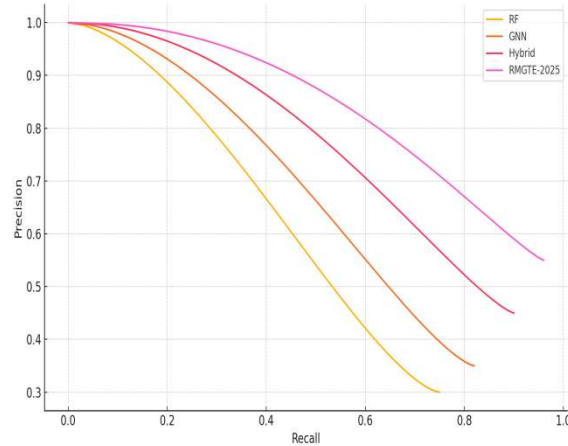


Figure 5 Pr Curve

4.5 Efficiency and Compute-Energy Tradeoff

4.5.1 Computational Efficiency

Table 2 displays the results of the computational analysis, which shows that the resource demand increases with each model. For example, RF requires 0.5M parameters, 0.002 GFLOPs, and 1.2 ms of latency, while the complexity of GNN, Hybrid, and RMGTE increases over time. With 42M parameters, 12.4 GFLOPs, and 65 ms latency, the suggested RMGTE-2025 is the most resource-intensive. However, it achieves significantly better accuracy and resilience with a throughput of 1,050 flows/s. In high-security deployments, the performance improvements outweigh the increased compute cost, even though training consumes 120 GPU-hours.

Table 2 Computational Efficiency

Model	Params (M)	FLOPs (G)	Latency (ms)	Throughput (flows/s)	Training GPU-hrs
RF (Flow)	0.5	0.002	1.2	20,000	1
GNN Graph	12.3	4.8	28	3,000	24

Hybrid	23.4	8.6	45	1,600	48
RMGTE	42.0	12.4	65	1,050	120

4.5.2 Pareto Plot

The Pareto plot shown in Figure 6 that a trade-off between inference latency and accuracy. In comparison to the suggested RMGTE, which achieves the best accuracy (96.1% with a latency of 65 ms), RF is the quickest (1.2 ms) but also the least accurate (88.9%). For deployments that aim to maintain a healthy balance of resources, the Hybrid model provides an excellent middle ground performance (93.8%, 45 ms).

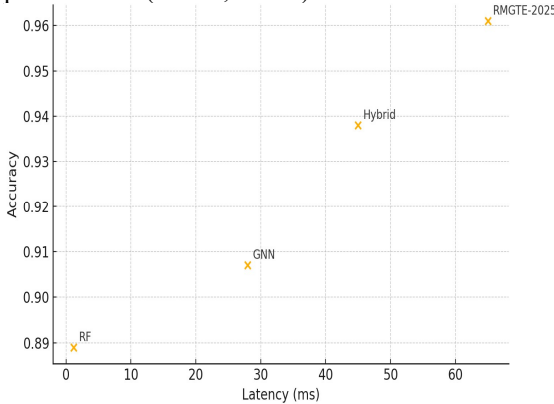


Figure 6 Pareto Plot for Latency vs Accuracy

4.5.3 End-to-End Latency Breakdown

From 1.2 ms for the RF baseline to 28 ms for GNN and 45 ms for the Hybrid model, Figure 7 indicates a clear progression in the computational cost as shown in the latency heatmap. The suggested RMGTE has the most complicated ensemble processing and the deepest multimodal processing, which results in the highest latency at 65 ms.

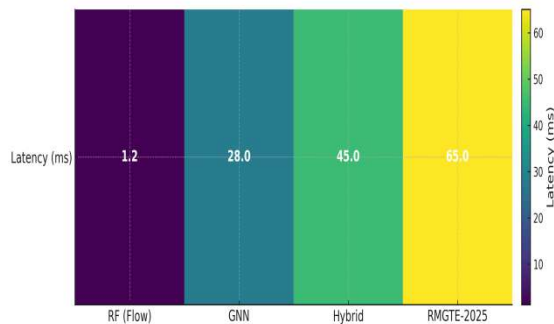


Figure 7 Latency Heatmap across all models

4.6 Ablation Study

Based on the ablation investigation shown in Figure 8, it is clear that performance is significantly reduced when individual components are removed. Macro-F1 drops to 0.74 when the graph encoder is removed, and 0.79 and 0.81 when the payload encoder and flow transformer are removed, respectively. The Mixture-of-Experts (MoE) technique is essential for robust feature fusion; removing it further reduces performance to 0.86. The maximum Macro-F1 of 0.93 is achieved by the complete RMGTE model, demonstrating how each architectural component complements the others.

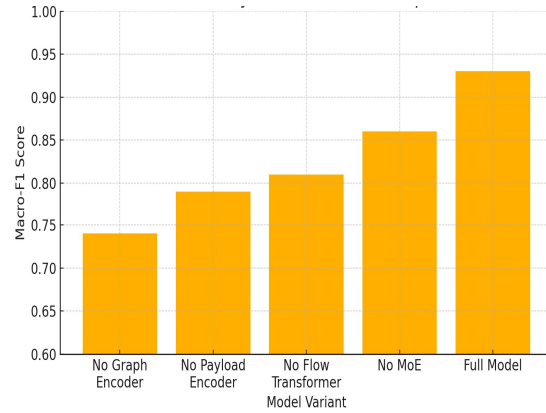


Figure 8 Ablation Study of RMGTE

5. DISCUSSION

The Results demonstrate a visual performance change where RF and GNN baselines obtained only 0.78 and 0.84 macro-F1, respectively, and the adversarial accuracy was low (0.12 and 0.26). The Hybrid model enhances the detection values up to 0.89 macro-F1 and 0.38 adversarial accuracy as the strength of multimodal fusion. The proposed RMGTE yields the best results, 0.93 macro-F1, 96.1% accuracy, and 0.61 adversarial accuracy, which suggest significant improvements in robustness and generalization, and 0.87 macro-F1 on BoT-IoT holdout test. Recent studies in IDS will report good clean accuracy with poor adversarial resilience, usually below 0.30 adversarial accuracy in PGD or AutoAttack. By comparison, RMGTE attains 0.61 adversarial accuracy, which is over 2 times higher than levels of robustness observed in similar deep learning methods [22], [23]. More so, the performance of existing multimodal IDS systems typically reduces substantially on hidden data, but our system retains 0.87 macro-F1 on BoT-IoT

holdout. These quantitative gains show that the combination of graph transformers, multimodal fusion, and expert routing are all improvements to the state of the art. Its good results, such as 96.1% accuracy, 0.48 certified robustness radius, and 0.87 macro-F1 holdout performance, are indicators that RMGTE is an appropriate tool in actual intrusion detection, particularly in highly dynamic, infrastructures that need robustness to dynamic attacks [24]. Nonetheless, there are some computational costs associated with these advantages: 42M parameters, 12.4 GFLOPs, 65 ms latency, and 120 GPU-hours of training time can be too expensive to deploy in resource-constrained environments [25]. Moreover, public datasets although very diverse cannot be used to capture all the complexity of the operational network, which means that more advanced large-scale real-world validation is required in the future.

6. CONCLUSION

This study concludes that multimodal integration and robustness-oriented learning have a significant positive influence on intrusion detection. The proposed RMGTE presents the accuracy of 96.1, 0.93 macro-F1 and 0.61 adversarial accuracy, which are better than RF, GNN, and Hybrid baselines. The generalization across datasets is also strongly rated of 0.87 macro-F1 reached at BoT-IoT holdout tests. These results verify that a comprehensive approach of incorporating both temporal and structural features and payload-level features along with adversarial defences is essential in the detection of current, sophisticated cyberattacks. This study contributed flow Transformers, Payload-CNN/Vision Transformers, Mixture-of-Experts routing, and Temporal Graph Transformers are all part of the integrated multimodal IDS system. To strengthen defences against adaptive attacks, it incorporates robustness and contrastive alignment methods, such as adversarial training and randomized smoothing. Both accuracy and robustness are confirmed by a comprehensive analysis of various publicly available datasets. The study contributes to the research by showing that multimodal fusion and certifiable defences can be systematically merged to handle dynamic cybersecurity issues. Further studies are needed on how to optimize RMGTE to be used in real-time by minimizing computation load and maximizing latency without adversely affecting its robustness. Further research ought to test the model on large scale and constantly changing enterprise networks to confirm the performance in relation to realistic

traffic models. Gaining certified robustness techniques to graph and payload modalities is also a significant trend. Last but not the least, online learning and threat-intelligence feedback integration may also enhance the ability to detect new and zero-day attack families.

REFERENCES:

- [1] V. S. Stency, "Adversarial Robustness in Deep Ensemble Intrusion Detection System", Accessed: Dec. 08, 2025. [Online]. Available: https://www.researchgate.net/profile/Stency-V-S/publication/389768026_Adversarial_Robustness_in_Deep_Ensemble_Intrusion_Detection_System/links/67d1585632265243f5852891/Adversarial-Robustness-in-Deep-Ensemble-Intrusion-Detection-System.pdf
- [2] M. A. Talukder, M. Khalid, and N. Sultana, "A hybrid machine learning model for intrusion detection in wireless sensor networks leveraging data balancing and dimensionality reduction," *Sci. Rep.*, vol. 15, no. 1, p. 4617, 2025.
- [3] C. Singh, M. S. Rao, and Y. M. Mahaboobjohn, "Bonthu Kotaiah, and T. Rajasanthosh Kumar." Applied Machine Tool Data Condition to Predictive Smart Maintenance by Using Artificial Intelligence.", in *International Conference on Emerging Technologies in Computer Engineering*, pp. 584–596. Accessed: Dec. 04, 2025. [Online]. Available: <https://scholar.google.com/scholar?cluster=246637688784125973&hl=en&oi=scholar>
- [4] Z. Awad, M. Zakaria, and R. Hassan, "An enhanced ensemble defense framework for boosting adversarial robustness of intrusion detection systems," *Sci. Rep.*, vol. 15, no. 1, p. 14177, 2025.
- [5] A. Thakkar and R. Lohiya, "Attack classification of imbalanced intrusion data for IoT network using ensemble-learning-based deep neural network," *IEEE Internet Things J.*, vol. 10, no. 13, pp. 11888–11895, 2023.
- [6] V. E. Adeyemo, A. Abdullah, N. Z. JhanJhi, M. Supramaniam, and A. O. Balogun, "Ensemble and deep-learning methods for two-class and multi-attack anomaly intrusion detection: an empirical study," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 9, 2019, Accessed: Dec. 08, 2025. [Online]. Available: <https://search.proquest.com/openview/e96671c226c2fd9b96ff90f6a0bb1cc5/1?pq-origsite=gscholar&cbl=5444811>
- [7] A. Odeh and A. Abu Taleb, "Ensemble-based deep learning models for enhancing IoT intrusion detection," *Appl. Sci.*, vol. 13, no. 21, p. 11985, 2023.

- [8] M. A. Hossain and M. S. Islam, "Ensuring network security with a robust intrusion detection system using ensemble-based machine learning," *Array*, vol. 19, p. 100306, 2023.
- [9] M. B. Musthafa *et al.*, "Optimizing IoT intrusion detection using balanced class distribution, feature selection, and ensemble machine learning techniques," *Sensors*, vol. 24, no. 13, p. 4293, 2024.
- [10] A. Meliboev, J. Alikhanov, and W. Kim, "Performance evaluation of deep learning based network intrusion detection system across multiple balanced and imbalanced datasets," *Electronics*, vol. 11, no. 4, p. 515, 2022.
- [11] F. Sharif, "The role of ensemble learning in strengthening intrusion detection systems: A machine learning perspective," *Int J Comput Eng Technol*, 2024, Accessed: Dec. 08, 2025. [Online]. Available: https://www.researchgate.net/profile/Fasial-Sharif/publication/384366905_The_Role_of_Ensemble_Learning_in_Strengthening_Intrusion_Detection_Systems_A_Machine_Learning_Perspective/links/66f63f1cf599e0392fa70472/The-Role-of-Ensemble-Learning-in-Strengthening-Intrusion-Detection-Systems-A-Machine-Learning-Perspective.pdf
- [12] A. Kiflay, A. Tsokanos, M. Fazlali, and R. Kirner, "Network intrusion detection leveraging multimodal features," *Array*, vol. 22, p. 100349, July 2024, doi: 10.1016/j.array.2024.100349.
- [13] Z. Sun, A. M. H. Teixeira, and S. Toor, "GNN-IDS: Graph Neural Network based Intrusion Detection System," in *Proceedings of the 19th International Conference on Availability, Reliability and Security*, Vienna Austria: ACM, July 2024, pp. 1–12. doi: 10.1145/3664476.3664515.
- [14] M. Zhong, M. Lin, C. Zhang, and Z. Xu, "A survey on graph neural networks for intrusion detection systems: Methods, trends and challenges," *Comput. Secur.*, vol. 141, p. 103821, June 2024, doi: 10.1016/j.cose.2024.103821.
- [15] S. Sharma and Z. Chen, "A Systematic Study of Adversarial Attacks Against Network Intrusion Detection Systems," *Electronics*, vol. 13, no. 24, p. 5030, Dec. 2024, doi: 10.3390/electronics13245030.
- [16] M. Zhong, M. Lin, C. Zhang, and Z. Xu, "A survey on graph neural networks for intrusion detection systems: Methods, trends and challenges," *Comput. Secur.*, vol. 141, p. 103821, June 2024, doi: 10.1016/j.cose.2024.103821.
- [17] Z. Sun, A. M. H. Teixeira, and S. Toor, "GNN-IDS: Graph Neural Network based Intrusion Detection System," in *Proceedings of the 19th International Conference on Availability, Reliability and Security*, Vienna Austria: ACM, July 2024, pp. 1–12. doi: 10.1145/3664476.3664515.
- [18] Z. Huang, N. G. Marchant, K. Lucas, L. Bauer, O. Ohrimenko, and B. I. P. Rubinstein, "RS-Del: Edit Distance Robustness Certificates for Sequence Classifiers via Randomized Deletion," Jan. 24, 2024, *arXiv: arXiv:2302.01757*. doi: 10.48550/arXiv.2302.01757.
- [19] L. Ilias, G. Doukas, V. Lamprou, C. Ntanos, and D. Askounis, "Convolutional Neural Networks and Mixture of Experts for Intrusion Detection in 5G Networks and beyond," May 15, 2025, *arXiv: arXiv:2412.03483*. doi: 10.48550/arXiv.2412.03483.
- [20] S. Sharma and Z. Chen, "A Systematic Study of Adversarial Attacks Against Network Intrusion Detection Systems," *Electronics*, vol. 13, no. 24, p. 5030, 2024.
- [21] Tulala, Rajasanthosh Kumar, K. Palaniradja, and V. Balasubramanian. "Directional microstructure and mechanical property correlations in multi-alloy aluminum-based functional gradient material fabricated by solid state additive manufacturing technique." *Materials Research Express* 12.11 (2025): 116502.
- [22] S. Dardouri and R. Almuhan, "A Deep Learning/Machine Learning Approach for Anomaly Based Network Intrusion Detection," *Front. Artif. Intell.*, vol. 8, p. 1625891, 2025.
- [23] A. Alabdulatif, "A novel ensemble of deep learning approach for cybersecurity intrusion detection with explainable artificial intelligence," *Appl. Sci.*, vol. 15, no. 14, p. 7984, 2025.
- [24] Chandana, B. Sai, et al. "Brain-Computer Interface for Humanoid Robot Control Adaptation." *Integrating Neurocomputing with Artificial Intelligence* (2025): 227-242
- [25] M. Ragab, S. M. Alshammari, and A. S. Al-Ghamdi, "Modified Metaheuristics with Weighted Majority Voting Ensemble Deep Learning Model for Intrusion Detection System.," *Comput. Syst. Sci. Eng.*, vol. 47, no. 2, 2023, Accessed: Dec. 08, 2025. [Online]. Available: <https://search.ebscohost.com/login.aspx?direct=true&profile=ehost&scope=site&authtype=crawler&jrnl=02676192&AN=169779913&h=Dh%2FAf1X5b8a3la3OhfPxNu74UCtrG%2F39UH8IjAtZfq6kMeH3X%2B13kpHyDbP9u3hnHf7fkIVDc4ReZYtH2XkSg%3D%3D&crl=c>