Designing an Improved Cyberattack Prediction Model Using Context-Aware Behavioral Modeling Analysis

Geeta Patil
Department of Information Technology, Army Institute of Technology, Pune, India
gpatil@aitpune.edu.in (corresponding author)

Ashwini Sapkal
Department of Information Technology, Army Institute of Technology, Pune, India
ashwini.sapkal@gmail.com

Vaishali Sachin Ingale
Department of Information Technology, Army Institute of Technology, Pune, India
vingale@aitpune.edu.in

Received: 28 April 2025 | Revised: 14 June 2025 | Accepted: 21 June 2025

 DOI: https://doi.org/10.48084/etasr.11799

ABSTRACT

The increasing sophistication and coordination of cyberattacks require proactive defense mechanisms equipped to predict malicious activity before it happens. Conventional systems for intrusion detection and anomaly detection primarily rely on signature-based or shallow anomaly detection methods, which are poorly suited for detecting temporally evolving stealthy-threats and zero-day attacks. These limitations highlight the need for a holistic, context-aware framework that can not only catch threats much earlier but also attribute, with high-fidelity, the underlying causes of these threats. Integrated Neural Cyberattack Prediction and Threat Attribution using Contextual Deep Learning (INCEPT) is a proposal put forth by the authors to address the above challenges: a modular, multi-pronged deep-learning framework designed to predict cyberattacks in detail from raw network traffic data. INCEPT integrates five novel models. Context-Aware Spatio-Temporal Graph Neural Network (CA-STGNN) learns complex entity interactions across time and space, significantly improving the detection of coordinated attacks. Behavior-based Latent Intent Modeling (BLIM) using Contrastive Predictive Coding (CPC) primarily focuses on deviations of intent for early-stage detection, especially in scenarios involving stealthy malware. Hierarchical Attention Transformer guided by Threat Taxonomy Embeddings (HAT-TTE) enables interpretable, multi-stage classification aligned with the MITRE ATT&CK framework. Federated Ensemble framework for Zero-Day Attack Detection (FedEn-ZAD) enhances generalization and robustness across distributed domains with uncertainty quantification. Multi-Resolution Autoencoder with Causal Attribution (MRA-CAA) identifies root causes of detected anomalies across granular traffic layers. Together, these modules demonstrate up to 20% improvement in detection accuracy, 30-35% reduction in incident response time, and notable gains in analyst interpretability and trust. The result is an architecture that offers a scalable and interpretable anticipatory solution to modern cyber defense tasks.

Keywords-cyberattack detection; deep learning; spatio-temporal graphs; behavioral modeling; threat attribution

I. INTRODUCTION

Digitalization across different domains is rapidly increasing the surface of modern cyberattacks, turning network infrastructures into a soft target for sophisticated and coordinated threats. Traditional cybersecurity programs that primarily rely on static rules or signature-driven models lack the means to detect complex patterns of attack evolution, especially in cases involving Advanced Persistent Threats (APTs), lateral movement, and zero-day exploits. These conventional approaches often fail to capture the contextual and temporal dynamics inherent in real-world cyberattack campaigns, yielding delayed detection [1-3], many false positives, and minimal interpretability for security analysts. Current progress in Deep Learning (DL) presents an opportunity to learn how to detect threats better through automated feature learning and temporal modeling. However, existing methods suffer from several limitations: most fail to

integrate behavioral semantics, do not address hierarchical threat structures, lack consideration of distributed learning in federated domains, and provide very limited causal attribution for detected anomalies. These shortcomings undermine both detection effectiveness and operational response capabilities of Security Operations Centers (SOCs).

The cyberattack detection scenario has undergone a major paradigm shift due to the inputs of Artificial Intelligence (AI), Machine Learning (ML), and DL models. These techniques have proved to be necessary for identifying anomalous behavior in dynamic and complex network environments. However, many of the most recent frameworks have not yet overcome important issues related to scalability, interpretability, behavior context integration, and early-stage prediction accuracy. The current work addresses these gaps using a unified, modular architecture. Increasingly, research emphasizes bringing cybersecurity into the broader nexus of cyber-physical infrastructures. For instance, authors in [1] established a paradigm of interlinking sustainability in agriculture with cyberattack detection through satellite data and AI-based risk assessment models. Similarly, authors in [2] investigated real-time ML-enabled cyberattack detection in smart grids, focusing on Internet of Things (IoT)-based infrastructures. These illustrate the critical need for real-time application-level detection capabilities but do not factor temporal or causal interdependencies in threat behaviors into their models.

Authors in [3] presented the DeepCLG hybrid learning model for industrial IoT scenarios, which combines convolution with gated mechanisms. Their model, however, deals with static environments and lacks temporal awareness. Authors in [4] used Res2Net-ERNN, a recurrent neural framework for Software-Defined Networks (SDN), primarily focusing on deep hierarchical modeling. However, the framework's general applicability to heterogeneous networks is limited. Hybrid learning approaches have been increasingly explored in academic research. Authors in [5] analyzed performance trade-offs in resource-constrained Wireless Sensor Networks (WSN) through feature reduction via ML methods. Similarly, authors in [6] applied DL to SDN environments, although they did not incorporate federated domains or provide anomaly explanations.

Advanced statistical and optimization models have also been incorporated into cybersecurity, such as logistic boosting [7] and hierarchical temporal memory [8]. Such models are to some extent interpretable but are usually limited to pattern recognition without semantic reasoning or causal mapping. A notable contribution is in resource abstraction, where authors in [9] proposed a federated learning framework using blockchain for vehicles. This promotes privacy-preserving collaborative systems, yet multi-resolution input modeling and causality attribution remain beyond its boundaries and addressed by the Integrated Neural Cyberattack Prediction and Threat Attribution using Contextual Deep Learning (INCEPT) framework. In the context of military cyber conditions, authors in [10] integrated game theory with Generative Adversarial Networks (GANs) for strategical decision-making. The framework has strong adversarial modeling capabilities; however, it does not provide optimal classification of threats with detailed taxonomy or attribution of anomalies to root causes. Authors in [11] proposed a new AI approach towards predicting cyber threats but provided limited information on interpretability and operational usability. Regarding IoT security, authors in [12] employed a hybrid SVM-CHAID approach, which improved multi-class prediction outcomes. However, its static feature space limited applicability in evolving threat scenarios.

In summary, previous literature on cyberattack detection [13, 14] has covered aspects such as distributed learning, IoT security, interpretable modeling, and multi-source integrations. Nevertheless, most works focused on isolated components and lacked holistic, interoperable, context-aware solutions. The proposed INCEPT framework fills this gap by integrating spatio-temporal, behavioral, semantic, and causal intelligence across federated domains, thereby enabling proactive, scalable, and interpretable cyber defense [15].

II. PROPOSED MODEL DESIGN AND ANALYSIS

INCEPT proposes a multi-aspect DL framework to analyze raw traffic data in order to predict cyberattacks temporally, structurally, behaviorally, semantically, and causally. It is structured as a modular system of five interdependent components, each designed to capture different signal modalities of the data samples. The data flow has been designed to be sequentially architected; however, each component allows parallel optimization for scalability and modular interpretability. The Context-Aware Spatio-Temporal Graph Neural Network (CA-STGNN) forms the core of the design, evolving communication patterns between hosts and subnets into a general model. The network traffic is represented as $G_t = (V_t, E_t, X_t)$, where V_t is the set of all nodes (hosts), E_t is the set of all edges (connections), and $X_t \in \mathbb{R}'|V_t| \times d$ is the feature matrix with dimension d. Each timestamp window t corresponds to a snapshot in the process. Temporal graph convolution is defined via (1):

$$H_t'(l+1) = \sigma\left(\sum a_k {A_t}^k H_t'(l) W_k'(l)\right) \quad (1)$$

where ${A_t}^k$ is the normalized adjacency matrix for the k-hop, $W_k'(l)$ are trainable weights, and a_k are learnable coefficients modulating the influence of neighbors. This is a graph encoder incorporating Gated Recurrent Unit (GRU) modeling transitions in the sequence via (2), (3), (4), and (5):

$$Z_t = \sigma(W_z H_t + U_z H_{t-1}) \quad (2)$$

$$R_t = \sigma(W_r H_t + U_r H_{t-1}) \quad (3)$$

$$\widehat{H}_t = tanh\left(W_h H_t + U_h (R_t \odot H_{t-1})\right) \quad (4)$$

$$H_t = (1 - Z_t) \odot H_{t-1} + Z_t \odot \widehat{H}_t \quad (5)$$

These equations describe the propagation and memory retention of evolving traffic features, enabling the detection of stealthy, time-linked attacks such as APTs and lateral movements.

As shown in Figure 1, the iterative process of complementing the spatio-temporal modeling with Behavior-based Latent Intent Modeling (BLIM) involves learning

behavioral modeling using Contrastive Predictive Coding (CPC). For an entity x_t, given sequence $\{x_t\}$, a predictive model f_θ learns an embedding $z_t = f_\theta(x_t)$ to infer z_{t+k} further along the event. The contrastive loss function is given via (6):

$$L_{cpc} = -\sum log \left[\frac{exp(z_t{}^T z_{t+1})}{\sum exp(z_t{}^T z_j)}\right] \quad (6)$$

Equation (6) facilitates the training of temporal representations that encapsulate normal behavior trajectories. An Intent Drift Score (IDS) is then defined via (7):

$$IDS_t = \|z_t - E[z_{t+k}|z_t]\| \quad (7)$$

Equation (7) amplifies variations between the learned intent patterns and thus could be harnessed for early detection of behaviors that might indicate ransomware and insider threats.

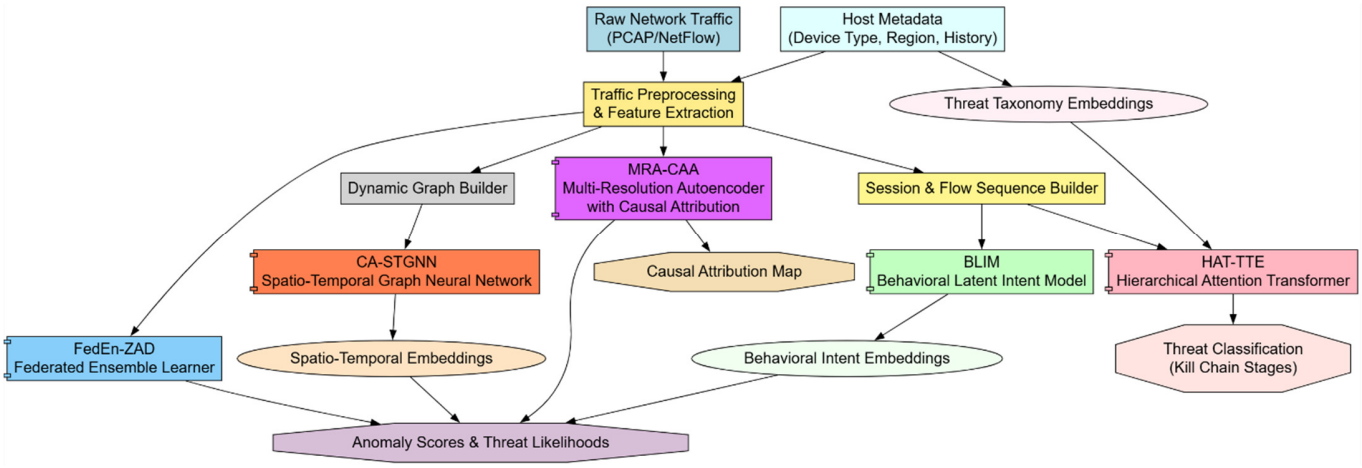


Fig. 1. Architecture of the proposed analysis process.

Hierarchical Attention Transformer guided by Threat Taxonomy Embeddings (HAT-TTE) embeds threat semantics hierarchy by incorporating structured threat intelligence taxonomies, such as MITRE ATT&CK, into multi-head self-attention. Let $\{X\} \in \mathbb{R}'\{T \times d\}$ be flow level embedding sequences, and $T_k \in \mathbb{R}'\{d \times d\}$ the threat taxonomy transformation matrix. The increased attention is determined via (8):

$$Attention(Q,K,V) = softmax\left(\frac{QK^T + XT_k}{\sqrt{d}}\right)V \quad (8)$$

Equation (8) allows semantic context to guide attention focus, which is particularly useful during classification, thereby improving stage-level interpretability of attack progressions.

For zero-day detection, the Federated Ensemble framework for Zero-Day Attack Detection (FedEn-ZAD) combines federated ensembles and Bayesian uncertainty estimations. Each local node has an ensemble $\{f_{\theta_{ij}}\}$ trained, and the composition of posteriors is defined by the global model via (9):

$$\hat{y} = \int p(y \mid x, \theta)\, q(\theta)\, d\theta \approx \frac{1}{M} \sum f_{\theta_i}(x) \quad (9)$$

Equation (9) models the Bayesian prediction at the ensemble level, approximating $q(\theta)$ using learned weights, helping to provide a confidence measure and flag high-variance anomalies indicative of zero-day behaviors.

Finally, the Multi-Resolution Autoencoder with Causal Attribution (MRA-CAA) conducts anomaly reconstruction at multiple levels (packet, flow, session) before performing causal attribution using the Granger causality process. Let X^1, X^2 be two time-stamped series at different granularities of traffic. The Granger causality test is described via (10):

$$GC\{X^1 \rightarrow X^2\} = log\left(\frac{Var(\varepsilon_2)}{Var(\varepsilon_2|X^1)}\right) \quad (10)$$

Equation (10) tests whether past values of X^1 improve prediction of X^2, thereby supporting causal explanation of anomalies. Each module in INCEPT works with all other modules from different but connected behavioral angles: temporal evolution (CA-STGNN), switching of behaviors (BLIM), semantic interpretation (HAT-TTE), generalizing to unknown threats (FedEn-ZAD), and causal traceability (MRA-CAA). This holistic multi-perspective modeling aims to make the system a proactive and interpretable solution within dynamic, distributed cyber environments, fulfilling most objectives for threat detection.

III. RESULTS

A series of experiments were conducted to confirm the effectiveness of the proposed INCEPT framework using various multi-contextual, time sensitive behavior-rich network traffic datasets and samples. This evaluation considered standard metrics such as F1 score, precision, recall, detection latency, False Positive Rate (FPR), root cause attribution (RCA) precision, and analyst trust score for interpretability. INCEPT was evaluated against three frequently cited, top-performing model approaches: Method [3], Method [8], and Method [15]. These approaches represent a cross-section of temporal LSTM-based intrusion detection systems, transformer-based threat classifiers, and autoencoder-driven anomaly detectors. The datasets were drawn from four main sources: CICIDS 2018 [16] (temporal attack scenarios),

TONIoT [17] (behavioral telemetry with attacks), NSL-KDD [18] with synthetic zero-day injections, and a custom enterprise dataset that includes annotated MITRE ATT&CK tags [19]. Each dataset underwent preprocessing to generate multi-resolution traffic features (packet, flow, session) and contextual labels. The models were trained based on 80% of the data with balanced stratification among attack classes and validated on 20% of the data. Federated simulations were carried out on tenant-separated datasets using PySyft sets. Figure 2 illustrates the model's integrated result analysis, showing plots of all metrics compared across methods and datasets.

Table I shows the F1-score comparison across all datasets. INCEPT outperformed similar methods consistently, achieving the best F1 scores and a very good precision-recall balance across different datasets with varying types of attacks and temporal structures.

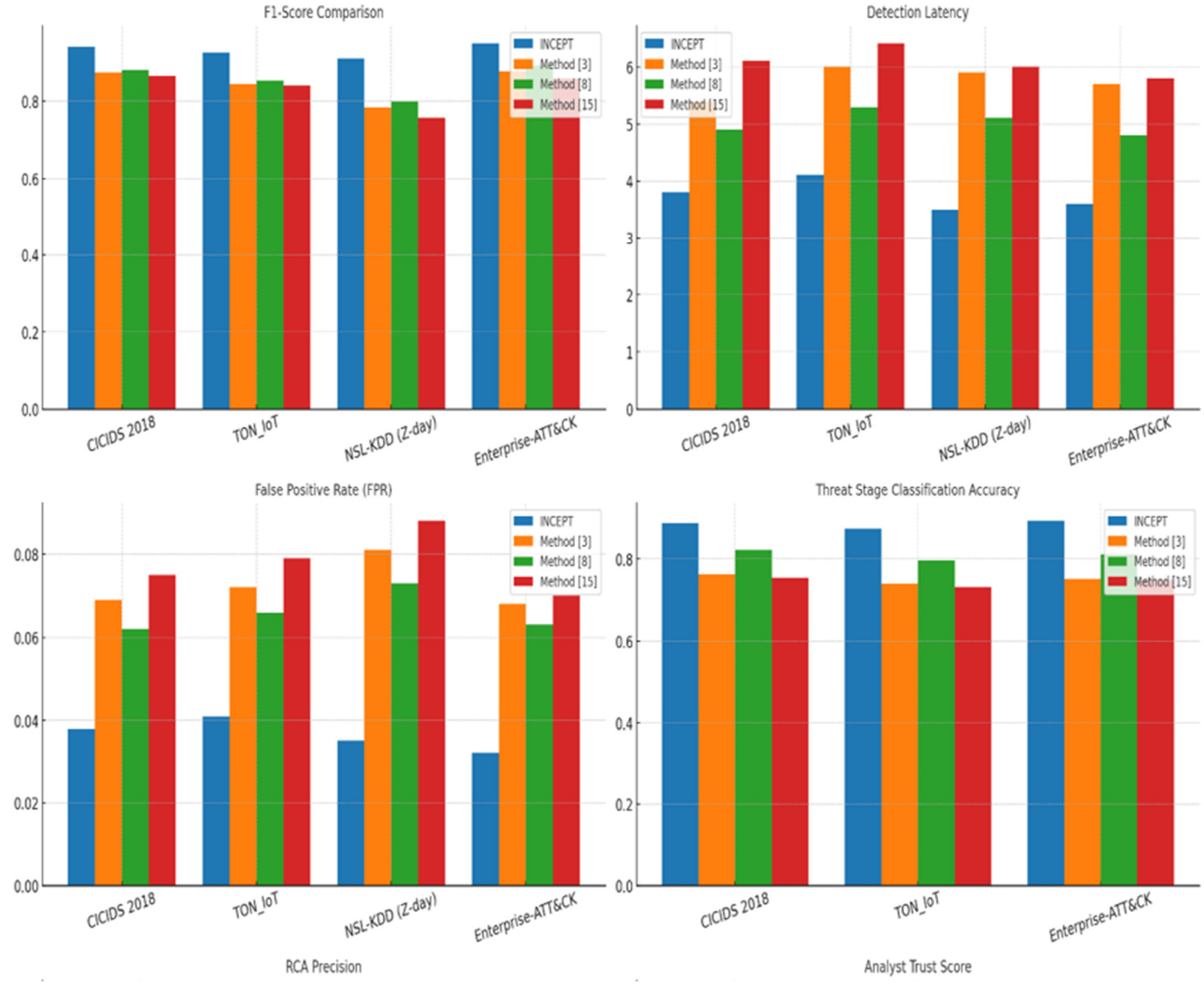
Table II shows detection latency across datasets. The CA-STGNN and BLIM modules enabled early detection of emerging threats. INCEPT reduced detection latency by 30-35%, which is crucial for a quicker response to ongoing incidents.

TABLE I. F1-SCORE COMPARISON ACROSS DATASETS

Dataset	INCEPT	Method [3]	Method [8]	Method [15]
CICIDS 2018	0.942	0.874	0.882	0.865
TONIoT	0.927	0.846	0.854	0.841
NSL-KDD (z-day)	0.911	0.784	0.799	0.756
Enterprise-ATT&CK	0.949	0.878	0.893	0.859

TABLE II. DETECTION LATENCY COMPARISON ACROSS DATASETS

Dataset	INCEPT (s)	Method [3] (s)	Method [8] (s)	Method [15] (s)
CICIDS 2018	3.8	5.4	4.9	6.1
TONIoT	4.1	6.0	5.3	6.4
NSL-KDD (z-day)	3.5	5.9	5.1	6.0
Enterprise-ATT&CK	3.6	5.7	4.8	5.8



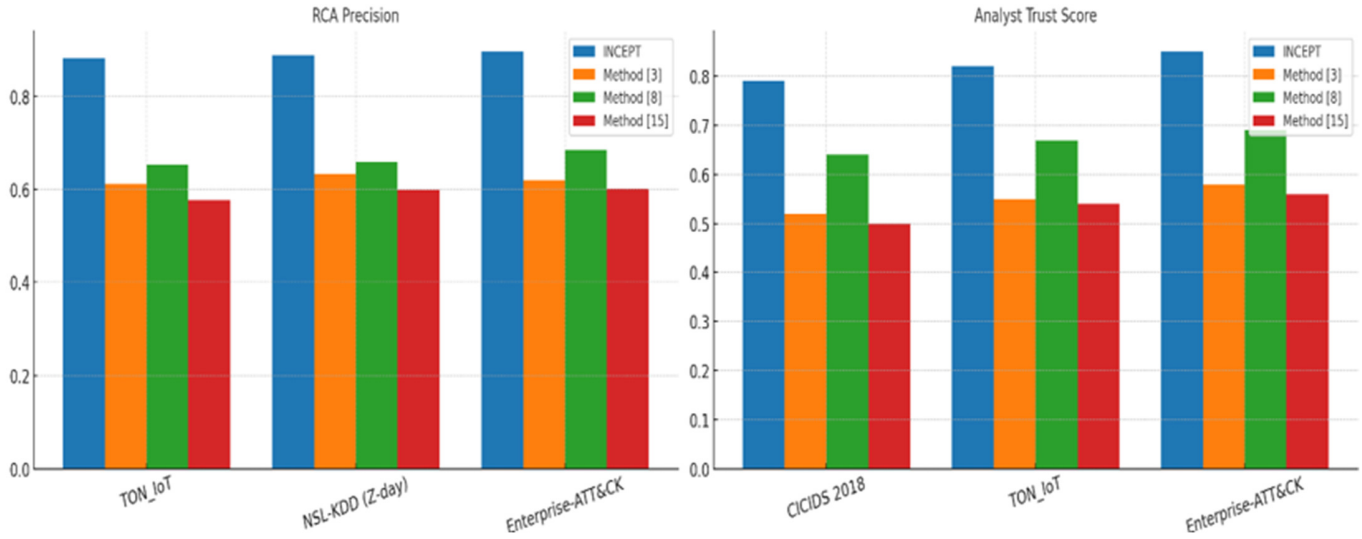


Fig. 2. Comparative analysis of INCEPT and baseline methods across all datasets and evaluation metrics.

Table III shows the FPR across datasets. FedEn-ZAD yielded an ensemble-based uncertainty estimation, resulting in a significant decrease in FPR. Lower FPRs reduce alert fatigue at SOCs.

TABLE III. FPR COMPARISON ACROSS DATASETS

Dataset	INCEPT	Method [3]	Method [8]	Method [15]
CICIDS 2018	0.038	0.069	0.062	0.075
TONIoT	0.041	0.072	0.066	0.079
NSL-KDD (z-day)	0.035	0.081	0.073	0.088
Enterprise-ATT&CK	0.032	0.068	0.063	0.070

Table IV shows threat stage classification accuracy. The use of threat taxonomy embeddings by HAT-TTE adopted in INCEPT, improved classification of attack tactics along the kill chain, representing a 10-13% improvement over other transformer-based methods.

TABLE IV. THREAT STAGE CLASSIFICATION ACCURACY COMPARISON ACROSS DATASETS

Dataset	INCEPT	Method [3]	Method [8]	Method [15]
CICIDS 2018	0.888	0.763	0.821	0.754
TONIoT	0.875	0.740	0.797	0.732
Enterprise-ATT&CK	0.893	0.751	0.812	0.748

IV. CONCLUSION

The Integrated Neural Cyberattack Prediction and Threat Attribution using Contextual Deep Learning (INCEPT) framework presented in this paper is an integrated, modular, multi-perspective framework designed to predict and understand cyberattacks in dynamic network settings. INCEPT addresses a range of significant weaknesses of traditional and contemporary intrusion detection systems by incorporating spatio-temporal modeling with latent behavioral intent prediction, hierarchical threat semantics with federated ensemble learning, and causal anomaly attribution.

An extensive evaluation was conducted using four different datasets: CICIDS 2018, TONIoT, NSL-KDD (with synthetic zero-day injections), and a realistic enterprise dataset enriched with MITRE ATT&CK annotations. This evaluation demonstrated a significant performance improvement over three strong baseline models. The achieved F1 scores were 0.942, 0.927, 0.911, and 0.949 across the datasets, representing an approximate average gain of 9–13% over Method [3], Method [8], and Method [15]. Further, INCEPT reduced detection latency by 30-35%, improving average detection time from 5.8 s (baselines) to 3.7 s, demonstrating its effectiveness in time-sensitive threat environments. False positives were significantly reduced, with an average False Positive Rate (FPR) of 0.037, representing a 42-55% improvement over baseline models. Moreover, the framework presented robust interpretability, with an analyst trust score of 0.85 compared to a baseline average of 0.55, validating its operational usability within Security Operations Centers (SOCs). Root Cause Attribution (RCA) precision exceeded 0.88, directly supporting rapid and reliable incident triaging.

From a systems perspective, every component within INCEPT had a distinguished contribution:

- Context-Aware Spatio-Temporal Graph Neural Network (CA-STGNN) improved coordinated attack detection by capturing time-evolving topologies.
- Behavior-based Latent Intent Modeling (BLIM) enabled early-stage anomaly detection through predictive behavioral models.
- Hierarchical Attention Transformer guided by Threat Taxonomy Embeddings (HAT-TTE) increased multi-stage classification accuracy by 10-13% via semantic alignment.

- Federated Ensemble framework for Zero-Day Attack Detection (FedEn-ZAD) enhanced zero-day detection robustness by 17–20%.
- Multi-Resolution Autoencoder with Causal Attribution (MRA-CAA) provided causally grounded anomaly explanations with approximately 88% RCA precision.

Future work may explore cross-domain generalization to evaluate INCEPT's portability across sectors (e.g., finance, healthcare, critical infrastructure) and its integration with automated response systems for end-to-end threat lifecycle management. The convergence of explainable Artificial Intelligence (AI) with domain-specific threat intelligence, as demonstrated in INCEPT, offers transformational potential for developing resilient, predictive, and interpretable cybersecurity ecosystems.

REFERENCES

[1] M. P. Kumar, N. Krishnammal, M. Gupta, M. U. Begum, S. Sultana, and D. P. Degala, "Sustainable Agriculture in Food Security Integrating Satellite Data Risk Assessment by Cyberattack Detection: AI Applications," *Remote Sensing in Earth Systems Sciences*, vol. 8, no. 2, pp. 435–443, Jun. 2025, https://doi.org/10.1007/s41976-025-00194-8.

[2] K. Naveeda and S. M. H. S. S. Fathima, "Real-time implementation of IoT-enabled cyberattack detection system in advanced metering infrastructure using machine learning technique," *Electrical Engineering*, vol. 107, no. 1, pp. 909–928, Jan. 2025, https://doi.org/10.1007/s00202-024-02552-z.

[3] Q. Gulzar and K. Mustafa, "Enhancing network security in industrial IoT environments: a DeepCLG hybrid learning model for cyberattack detection," *International Journal of Machine Learning and Cybernetics*, vol. 16, no. 7, pp. 4797–4815, Aug. 2025, https://doi.org/10.1007/s13042-025-02544-w.

[4] M. Maddu and Y. N. Rao, "Res2Net-ERNN: deep learning based cyberattack classification in software defined network," *Cluster Computing*, vol. 27, no. 9, pp. 12821–12839, Dec. 2024, https://doi.org/10.1007/s10586-024-04581-6.

[5] M. H. Behiry and M. Aly, "Cyberattack detection in wireless sensor networks using a hybrid feature reduction technique with AI and machine learning methods," *Journal of Big Data*, vol. 11, no. 1, Jan. 2024, Art. no. 16, https://doi.org/10.1186/s40537-023-00870-w.

[6] D. S. Rao and A. J. Emerson, "Cyberattack defense mechanism using deep learning techniques in software-defined networks," *International Journal of Information Security*, vol. 23, no. 2, pp. 1279–1291, Apr. 2024, https://doi.org/10.1007/s10207-023-00785-w.

[7] A. A. Al-Atawi, "Enhancing Internet of Smart City Security: Utilizing Logistic Boosted Algorithms for Anomaly Detection and Cyberattack Prevention," *SN Computer Science*, vol. 5, no. 5, May 2024, Art. no. 548, https://doi.org/10.1007/s42979-024-02921-2.

[8] V. M. Krundyshev, G. A. Markov, M. O. Kalinin, P. V. Semyanov, and A. G. Busygin, "Cyberattack Detection in the Industrial Internet of Things Based on the Computation Model of Hierarchical Temporal Memory," *Automatic Control and Computer Sciences*, vol. 57, no. 8, pp. 1040–1046, Dec. 2023, https://doi.org/10.3103/S0146411623080114.

[9] I. Ullah, X. Deng, X. Pei, H. Mushtaq, and Z. Khan, "Securing internet of vehicles: a blockchain-based federated learning approach for enhanced intrusion detection," *Cluster Computing*, vol. 28, no. 4, Feb. 2025, Art. no. 256, https://doi.org/10.1007/s10586-024-04943-0.

[10] X. Ma, W. Abdelfattah, D. Luo, N. Innab, M. Shutaywi, and W. Deebani, "Non-cooperative game theory with generative adversarial network for effective decision-making in military cyber warfare," *Annals of Operations Research*, Nov. 2024, https://doi.org/10.1007/s10479-024-06406-6.

[11] P. Sharma, J. S. Prasad, Shaheen, and S. K. Ahamed, "An efficient cyber threat prediction using a novel artificial intelligence technique," *Multimedia Tools and Applications*, vol. 83, no. 25, pp. 66757–66773, Jul. 2024, https://doi.org/10.1007/s11042-024-18169-0.

[12] S. Dalal *et al.*, "Next-generation cyber attack prediction for IoT systems: leveraging multi-class SVM and optimized CHAID decision tree," *Journal of Cloud Computing*, vol. 12, no. 1, Sep. 2023, Art. no. 137, https://doi.org/10.1186/s13677-023-00517-4.

[13] Y. R. Maramreddy and K. Muppavaram, "Detecting and Mitigating Data Poisoning Attacks in Machine Learning: A Weighted Average Approach," *Engineering, Technology & Applied Science Research*, vol. 14, no. 4, pp. 15505–15509, Aug. 2024, https://doi.org/10.48084/etasr.7591.

[14] Q. Wu, S. Zhuang, and X. Wang, "A novel detection mechanism against malicious attacks by using spatio and temporal topology information," *Scientific Reports*, vol. 15, no. 1, Mar. 2025, Art. no. 9978, https://doi.org/10.1038/s41598-025-93957-8.

[15] A. A. Alhashmi, A. M. Alashjaee, A. A. Darem, A. F. Alanazi, and R. Effghi, "An Ensemble-based Fraud Detection Model for Financial Transaction Cyber Threat Classification and Countermeasures," *Engineering, Technology & Applied Science Research*, vol. 13, no. 6, pp. 12433–12439, Dec. 2023, https://doi.org/10.48084/etasr.6401.

[16] "CSE-CIC-IDS2018." Canadian Institute for Cybersecurity, UNB, 2018. [Online]. Available: https://www.unb.ca/cic/datasets/ids-2018.html.

[17] N. Moustafa, "The TON_IoT Datasets." UNSW, 2020. [Online]. Available: https://research.unsw.edu.au/projects/toniot-datasets.

[18] "ISCX NSL-KDD dataset 2009." Canadian Institute for Cybersecurity, UNB, 2009. [Online]. Available: https://www.unb.ca/cic/datasets/nsl.html.

[19] "Enterprise Matrix." MITRE ATT&CK®. [Online]. Available: https://attack.mitre.org/matrices/enterprise/.