Journal of Scientific Innovation and Advanced Research (JSIAR)
E-ISSN: 3107-507X
Volume 1, Issue 5, August 2025
https://jsiar.com

Advanced Graph Neural Network Techniques for Network Intrusion Detection: A Systematic Review and Future Directions

Jyoti Mahur
Department of Computer Science and Engineering
Noida International University, Greater Noida, India
Email: jyotimahur3oct@gmail.com

***Abstract*—The increasing sophistication of cyber threats and the complexity of modern network infrastructures have amplified the need for intelligent and adaptive intrusion detection systems (IDS). Traditional machine learning and deep learning methods often struggle to model the dynamic and relational characteristics inherent in network data. In this context, Graph Neural Networks (GNNs) have emerged as a powerful paradigm for learning structured representations from graph-based network traffic, enabling more accurate detection of malicious activities and anomalous behaviors. This paper presents a systematic review of advanced GNN techniques applied to network intrusion detection, encompassing architectural innovations, benchmark datasets, and performance trends reported across recent studies. The review follows a structured methodology, analyzing literature from 2018 to 2025 across major academic databases, and classifies existing approaches based on their graph modeling strategies, learning mechanisms, and detection objectives. Key findings indicate that GNN-based models significantly enhance detection precision, scalability, and resilience against evolving attack patterns. However, challenges remain in addressing explainability, computational efficiency, and real-time adaptability. The paper concludes by outlining future research directions, including the integration of explainable AI, federated learning frameworks, and hybrid GNN architectures to achieve interpretable, privacy-preserving, and adaptive intrusion detection solutions.**

***Keywords*—Graph Neural Networks (GNNs), Intrusion Detection Systems (IDS), Cybersecurity, Network Analysis, Deep Learning, Threat Detection, Explainable AI**

I. Introduction

In recent years, the volume, velocity, and variety of network traffic have grown exponentially, driven by the proliferation of connected devices, cloud services, and remote work. Consequently, modern cyber-threat landscapes have become increasingly complex, exhibiting sophisticated intrusion patterns and multi-stage attacks that evade legacy detection mechanisms. Traditional intrusion detection systems (IDS) that rely on rules or handcrafted signatures are increasingly inadequate in capturing novel or evolving threats, and machine learning (ML)-based methods have been introduced to improve adaptability and accuracy. However, even these ML techniques often treat each traffic sample in isolation, lacking the capability to fully exploit the relational and structural dependencies present in network flows.

Over the past decade, deep learning (DL) models such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs) have been applied for intrusion detection with improved results over classical ML techniques. Nevertheless, these models tend to operate on “flat” tabular or sequential data representations and thus may not capture the rich topological information inherent in communication networks. The emergence of graph-based approaches allows network entities (e.g., hosts, flows, packets) to be represented as nodes and relationships (e.g., communications, attacker-target links) as edges, thereby enabling a more faithful modelling of network behaviour. In this context, graph neural networks (GNNs) have emerged as a powerful paradigm for learning representations from graph-structured data, combining relational modelling with deep learning capacity [1]–[5], [7], [8], [12], [13].

The application of GNNs in cybersecurity—and specifically for network intrusion detection—has gained traction because of their ability to model interactions, capture global context, and learn robust embeddings from connectivity patterns. For example, recent surveys show that GNN-based IDS offer advantages in capturing structural behaviour of attacks and in improving robustness to adversarial perturbations [6], [9], [10], [14], [15], [33]. Moreover, in IoT and edge-driven environments where topological interactions matter, GNNs have been shown to outperform traditional ML and DL methods [11], [16], [17], [34], [35], [40]–[42], [44]. Despite these promising developments, the literature remains fragmented: there is no comprehensive review that systematically organises advances in GNN architectures, network intrusion datasets, performance trends, and deployment-level challenges in one place.

This paper seeks to fill that gap by presenting a systematic review of advanced GNN techniques for network intrusion detection. Specifically, our objectives are four-fold: (i) to map the evolution of GNN architectures applied to intrusion detection, (ii) to classify and compare the datasets, evaluation metrics, and application scenarios used in the field, (iii) to identify and analyse the major strengths, limitations, and gaps in current GNN-based intrusion detection research, and (iv) to propose future research directions that address emerging needs such as explainability, real-time deployment, and privacy-preserving learning. The contributions of this work are summarised as follows:

- We provide a detailed taxonomy of GNN models used in intrusion detection, including variants such as Graph Convolutional Networks (GCNs), Graph Attention Networks (GATs), GraphSAGE, and temporal GNNs.
- We present a comparative analysis of benchmark intrusion detection datasets (e.g., NSL-KDD, UNSW-NB15, CICIDS2017), model performance, and evaluation metrics.

TABLE I: Comparison between Traditional Deep Learning and GNN-based Intrusion Detection Systems

Aspect	**Traditional Deep Learning IDS (e.g., CNN, RNN)**	**GNN-based IDS**
Data Representation	Operates on tabular or sequential data; each sample independent	Models entities (hosts, flows) as graph nodes with relational edges
Structural Awareness	Limited to local spatial or temporal patterns	Captures topological and relational dependencies across network
Feature Engineering	Requires extensive preprocessing and manual feature selection	Learns relational embeddings directly from graph structure
Adaptability to Topology Changes	Poor adaptability to dynamic or evolving network graphs	Natively supports evolving graphs and incremental updates
Explainability	Hidden-layer features often opaque; difficult to interpret	Node- and edge-level importance can improve explainability
Scalability	Efficient on fixed-size data but struggles with high-dimensional network graphs	Computationally intensive for large graphs; requires sampling or clustering
Resilience to Adversarial Attacks	Susceptible to crafted perturbations	Improved resilience through relational reasoning and context aggregation
Application Domains	Traditional enterprise IDS, flow classification, anomaly detection	IoT, edge computing, blockchain, multi-agent and distributed systems

- We highlight key challenges in applying GNNs to network security, including scalability, explainability, dynamic topology adaptation, and adversarial vulnerabilities.
- We outline actionable future directions for academic researchers and practitioners, covering federated GNN frameworks, hybrid graph-temporal architectures, and standardised benchmarks for reproducibility.

The remainder of this paper is organised as follows. Section II presents the theoretical foundations of graph neural networks and the basics of intrusion detection systems. Section III details our systematic review protocol, including search strategy, inclusion/exclusion criteria, and data extraction process. Section IV introduces our taxonomy of GNN-based intrusion detection approaches. In Section V, we perform a comparative analysis of methods, datasets, and findings. Section VI discusses the key limitations and open issues in this area. Section VII outlines future research directions. Finally, Section VIII concludes the paper with summarised insights.

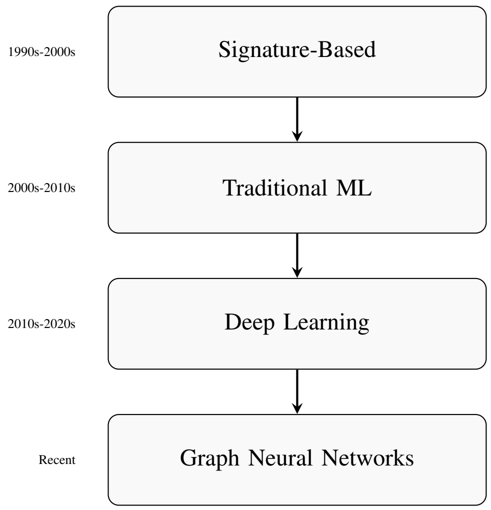

Fig. 1: Evolution of intrusion detection systems toward GNN-based approaches.

II. Background and Theoretical Foundations

A. Graph Theory in Network Modeling

Graph theory provides the foundational framework for representing and analysing network topologies, where entities such as hosts, flow records, or devices are modelled as nodes and relationships (e.g., communication links, session flows, attacker-target links) as edges. A typical network can be represented as $G = (V, E, A, X)$, where V is the set of nodes, $E \subseteq V \times V$ is the set of edges, A is the adjacency matrix capturing connectivity, and X represents node attributes. By exploiting these representations, one can capture multi-hop dependencies, propagation of malicious behaviour, and topological features such as centrality and community structure. In the context of intrusion detection, leveraging graph-based modelling enables the capture of structural context beyond isolated feature vectors, thereby providing a richer basis for detecting coordinated or stealthy attacks.

B. Fundamentals of Graph Neural Networks

Graph Neural Networks (GNNs) extend deep learning into the non-Euclidean domain of graphs, enabling representation learning over nodes, edges and entire graphs. Several core architectures have emerged:

- *Graph Convolutional Networks (GCNs)*: These apply spectral or spatial convolution operators on the graph structure to aggregate feature information from neighbourhoods. A typical propagation step can be expressed as
$$H^{(k+1)} = \sigma\left(\tilde{D}^{-1/2}\tilde{A}\tilde{D}^{-1/2}H^{(k)}W^{(k)}\right),$$
where $\tilde{A} = A + I$ and $H^{(k)}$ is the hidden representation at layer k. GCNs serve as a baseline variant of GNNs. [1]
- *Graph Attention Networks (GATs)*: GATs introduce attention coefficients α_{ij} to weight the contributions from different neighbours $j \in \mathcal{N}(i)$. This gives the propagation
$$h_i^{(k+1)} = \sigma\Big(\sum_{j \in \mathcal{N}(i)} \alpha_{ij} W h_j^{(k)}\Big),$$
where α_{ij} itself is computed using a learnable mechanism.

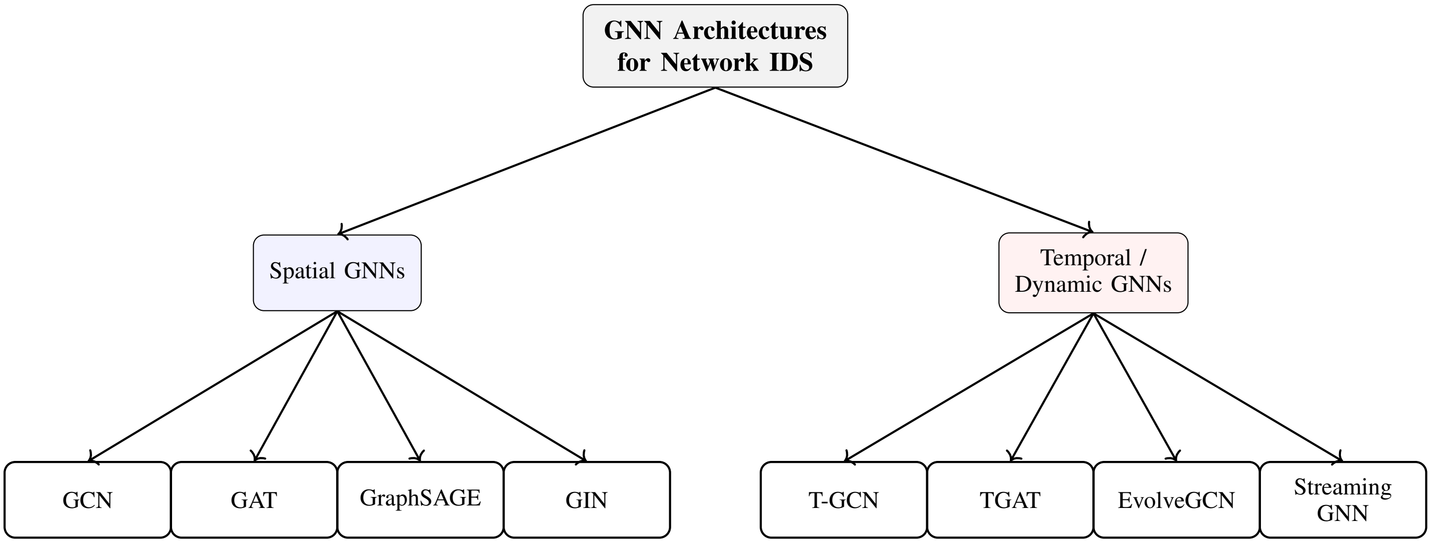

Fig. 2: Taxonomy of GNN architectures for network intrusion detection, highlighting the distinction between spatial and temporal approaches.

- *GraphSAGE (Sample and Aggregate)*: Designed for inductive learning on large graphs, GraphSAGE samples a fixed number of neighbours and aggregates their features, thereby enabling scalability and generalization to unseen nodes.
- *Graph Isomorphism Networks (GINs)*: GINs aim to remedy expressiveness limitations of previous GNNs by modifying the aggregation and read-out functions to achieve discriminatory power comparable to the Weisfeiler–Lehman graph isomorphism test.

These variants differ in their neighbourhood aggregation schemes, attention mechanisms, sampling strategies, and expressivity characteristics.

C. Spatial versus Temporal GNNs in Dynamic Network Analysis

In network intrusion detection, traffic flows evolve both in terms of network topology (spatial dimension) and attack sequence or time (temporal dimension). Thus, GNN models can be broadly categorised as:

- *Spatial GNNs*: These models focus on modelling the static or snapshot graph structure at a given time—aggregating across neighbours, capturing connectivity patterns or host-flow graphs.
- *Temporal (or Spatio-temporal) GNNs*: These extend spatial GNNs by incorporating time dynamics—either by stacking graph snapshots through time, leveraging recurrent modules (e.g., GNN+LSTM) or using temporal message passing. Such models are particularly suited for evolving network environments with dynamically changing topology or attack propagation over time.

Figure 2 provides a high-level taxonomy of GNN architectures used for network intrusion detection (spatial vs temporal).

D. Fundamentals of Network Intrusion Detection Systems (IDS)

An intrusion detection system (IDS) monitors network or host activities for suspicious behaviour and issues alerts. Broadly, there are two modes: signature-based detection (matching known attack patterns) and anomaly-based detection (learning normal behaviour and flagging deviations). In modern network environments, challenges include high dimensionality of features, class imbalance (few attacks vs many normal events), evolving or zero-day attacks, and real-time constraints. Traditional ML approaches (e.g., decision trees, SVMs) rely on flat tabular features, whereas deep learning approaches (e.g., CNNs/RNNs) model flow sequences or payloads but still often ignore the relational structure between hosts or flows.

E. Integration of GNNs into IDS Frameworks

The integration of GNNs into IDS frameworks involves several steps:

1) Graph Construction: Converting network flows or events into a graph representation (nodes = hosts/flows; edges = communications, similarity, temporal links).
2) Embedding/Representation Learning: Applying a GNN variant to learn node/edge/graph embeddings that capture structural and attribute information.
3) Detection or Classification Module: Utilizing the learned embeddings for downstream tasks (binary intrusion detection, multi-class attack classification, anomaly detection).
4) Deployment: Real-time or near-real-time inference, potentially on edge or distributed settings, considering scalability and update dynamics.

The recent works demonstrate that GNN-based IDS can outperform traditional ML/DL models in terms of detection accuracy and adaptability in complex network settings. For

TABLE II: Comparison of Core GNN Architectures

Architecture	Aggregation Scheme	Strengths	Limitations
GCN	Neighbour averaging with normalized adjacency	Simplicity, spectral interpretation	Requires full graph, less expressive
GAT	Attention-based neighbour weighting	Differentiates neighbour importance	Higher computation, memory cost
GraphSAGE	Neighbour sampling + aggregator	Scalability, inductive generalisation	Sampling bias, less expressive than full models
GIN	Sum aggregator + MLP read-out	High expressivity, theoretically powerful	More parameters, risk of overfitting

example, a device-level IIoT GNN-based IDS achieved very high accuracies by modelling both temporal and spatial dimensions. Nevertheless, challenges remain in graph construction, dynamic topology adaptation, explainability and real-time deployment.

F. Comparison of Core GNN Architectures

To facilitate understanding of various architectures, Table II summarises their key computational and modelling characteristics.

This section has established the theoretical foundations underpinning our review: the representation of networks through graphs, the evolution of GNN architectures and their spatial/temporal adaptations, fundamental IDS concepts, and how GNNs integrate into IDS workflows. In subsequent sections, we will build on these foundations to present our systematic review methodology, taxonomy of GNN-based intrusion detection approaches, and comparative analysis of recent research.

III. Research Methodology for Systematic Review

A systematic review aims to identify, evaluate, and synthesise the existing body of knowledge in a structured, reproducible manner. In this study, we adopted a transparent and replicable protocol inspired by the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines to ensure methodological rigour and objectivity. The following subsections detail the literature search process, inclusion and exclusion criteria, review protocol, and data extraction strategy used in this work.

A. Literature Search Strategy

The literature search was conducted across several major scientific databases to ensure comprehensive coverage of the research landscape. Specifically, the databases queried include IEEE Xplore, SpringerLink, ScienceDirect, ACM Digital Library, Scopus, and Wiley Online Library. The search terms were constructed using Boolean operators and relevant keywords such as *"graph neural network"*, *"intrusion detection"*, *"cybersecurity"*, *"network anomaly detection"*, and *"graph-based learning"*. An example search query is shown below:

> *Query:* ("graph neural network" OR "GNN" OR "graph learning") AND ("intrusion detection" OR "cyber attack" OR "network security")

The search period spanned publications from 2017 to 2025, aligning with the rapid advancement of GNN architectures during this period. Only peer-reviewed journal articles, conference proceedings, and reputable preprints (e.g., arXiv) were considered.

B. Inclusion and Exclusion Criteria

To ensure quality and relevance, inclusion and exclusion criteria were carefully designed, as summarised in Table III. Studies were included if they (i) proposed, implemented, or evaluated a GNN-based approach for intrusion detection or cybersecurity tasks, (ii) presented sufficient methodological detail for replication or comparison, and (iii) reported empirical results on recognised datasets. Exclusion criteria removed papers that were (i) not written in English, (ii) purely theoretical without evaluation, (iii) focused on non-network graph domains (e.g., social networks, molecules), or (iv) duplicated in multiple venues.

TABLE III: Inclusion and Exclusion Criteria for the Systematic Review

Inclusion Criteria	Exclusion Criteria
GNN applied to intrusion or anomaly detection	Non-cybersecurity or non-network domains
Peer-reviewed journal or conference publication	Unpublished theses, blogs, or reports
Contains experimental validation	Purely theoretical or conceptual frameworks
Uses standard datasets (e.g., NSL-KDD, CICIDS2017, UNSW-NB15)	Uses proprietary or unavailable datasets
English language publications (2017–2025)	Non-English or duplicate entries

C. Review Protocol

The review followed a PRISMA-style four-phase process: *Identification*, *Screening*, *Eligibility*, and *Inclusion*. During the identification phase, all retrieved publications were imported into reference management software (Zotero and Mendeley) for de-duplication. In the screening phase, titles, abstracts, and keywords were assessed for relevance. The eligibility phase involved full-text evaluation of potentially relevant papers, and finally, the inclusion phase selected studies that met all criteria.

D. Data Extraction Process

From each included study, a structured data extraction form was employed to capture both qualitative and quantitative characteristics. The key attributes extracted are listed below:

- *Bibliographic Information:* Authors, publication year, venue, and citation count.
- *Technical Features:* Type of GNN model (GCN, GAT, GraphSAGE, GIN, temporal GNN, etc.).
- *Dataset Information:* Benchmark datasets used (e.g., NSL-KDD, CICIDS2017, TON_IoT, UNSW-NB15).
- *Evaluation Metrics:* Accuracy, precision, recall, F1-score, AUC, and computational cost.

- *Architectural Insights:* Graph construction strategy, number of layers, attention mechanisms, sampling method.
- *Application Domain:* IoT, industrial networks, cloud computing, or hybrid architectures.
- *Key Findings:* Performance trends, strengths, limitations, and future challenges identified by the authors.

Data were manually extracted and cross-verified by two reviewers to ensure consistency and reduce subjective bias. Where applicable, data were normalised to ensure comparability across studies.

E. Bibliometric and Statistical Analysis

A bibliometric analysis was conducted to reveal trends in publication volume, model types, and dataset usage across time. Frequency distributions were computed for publication years, citation counts, and dataset adoption. Descriptive statistics (mean, median, standard deviation) were calculated to characterise performance metrics across reviewed studies. Additionally, network visualisation tools such as VOSviewer were used to analyse co-authorship networks, keyword co-occurrences, and citation clustering patterns. These analyses provided quantitative insights into the evolution of GNN-based intrusion detection research and identified underexplored areas ripe for further investigation.

This systematic review methodology ensures transparency, reproducibility, and comprehensive coverage of the current research landscape. By combining a PRISMA-inspired protocol, clearly defined selection criteria, and rigorous data extraction, this section lays the foundation for the forthcoming analysis and taxonomy presented in subsequent sections.

IV. Taxonomy of GNN-Based Intrusion Detection Approaches

In this section we present a structured taxonomy of intrusion detection systems that leverage graph neural network (GNN) models. Our classification is organised along three primary dimensions: (1) model architecture (e.g., Graph Convolutional Network (GCN), Graph Attention Network (GAT), GraphSAGE, Gated Graph Neural Network (GGNN), etc.), (2) application type (such as anomaly detection, malware detection, botnet analysis), and (3) data type (for instance network-flow graphs, host-communication graphs, attack-provenance graphs). This multi-facet taxonomy helps in mapping the literature, comparing approaches and identifying gaps. Leading recent reviews adopt a similar problem-oriented taxonomy for GNN-based IDS research. [30], [31]

A. Classification by Model Architecture

The choice of GNN architecture has significant implications for how the system captures structural dependencies, computational efficiency and interpretability. GCN-based models implement neighbourhood aggregation via spectral or spatial convolution operators and have been used in early GNN-IDS work given their simplicity and effectiveness. [32] GAT-based models introduce attention mechanisms on graph edges and nodes, enabling the model to weight neighbour contributions dynamically; these are increasingly used in intrusion detection for distinguishing contributing hosts or flows. [36] GraphSAGE and inductive models enable the handling of unseen nodes or streaming data by employing node sampling and aggregator functions; this is valuable in dynamic network traffic scenarios. [37] GGNN or recurrent/temporal GNN variants extend the spatial GNN to capture sequential behaviour or edge evolution over time, which is crucial for attack propagation modelling.

B. Classification by Application Type

GNN-based IDS studies can also be categorised by the attack or anomaly domain they target: - Anomaly detection: Detecting rare or unknown deviations in network graphs, often via unsupervised or self-supervised GNN embeddings. [38] - Malware detection and host compromise: Host-level graphs (system calls, process flows) modelled via GNNs to identify compromised machines or lateral movement. - Botnet and distributed attack detection: Graph structures representing device clusters, command-and-control channels or peer-to-peer botnet flows are analysed via GNNs to identify coordinated attacks. - IoT/edge network intrusion detection: Given the importance of device graphs and heterogeneous topologies, a growing number of works apply GNNs in IoT contexts leveraging communication graphs and edge constraints. [39]

C. Classification by Data Type

The third classification dimension is the form of graph data fed to the GNN: - Network-flow graphs model traffic flows as edges between hosts (nodes) and capture temporal or directional information. - Host-communication graphs represent host entities and their interaction patterns, often with heterogeneous node/edge types. - Attack-provenance graphs trace the lineage of events (e.g., system calls, user actions, network hops) into graphs for detection of multi-stage attacks. Each data type imposes specific design requirements: flow graphs emphasise high-volume streaming, host graphs emphasise heterogeneity and provenance graphs emphasise temporal and causality relationships. Prior surveys emphasise the importance of selecting the appropriate graph representation before applying a GNN model. [30]

D. Comparative Summary of Methods, Datasets and Performance

To facilitate comparison across approaches, Table IV lists representative GNN-IDS studies, the architecture used, the graph-type, dataset(s) employed, and reported performance.

E. Visual Taxonomy Diagram

Figure 3 presents a hierarchical diagram of this taxonomy, showing how model architectures, application domains and data types interrelate in GNN-based intrusion detection.

TABLE IV: Representative GNN-Based IDS Studies: Architectures, Data Types and Datasets

Study	**Architecture**	**Graph Data Type**	**Dataset(s)**	**Reported Metric**
Lo et al. (2021) E-GraphSAGE [45]	GraphSAGE with edge features	Network flow graph	NSL-KDD, CICIDS2017	Accuracy 95%
Tran	Park (2024) FN-GNN [37]	Hybrid GCN + GraphSAGE	Flow graphs with IP relations	CICIDS2017, UNSW-NB15 & Improved F1 over baselines
Zhong et al. (2024) Survey work [30]	Taxonomy	Mixed graph types	—	—
Springer BS-GAT (2024) [36]	GAT variant	Host/device communication graph	Edge-computing IoT dataset	Multi-class accuracy gain
Industrial IoT GIDS (2023) [39]	GNN (custom)	IoT device graph	BoT-IoT, ACI-IoT-2023	F1 > 94 %

F. Discussion and Research Gaps

Using the above taxonomy, several observations emerge. While many studies focus on flow-graph modelling and GCN/GAT architectures, fewer works address provenance graph modelling or temporal graph dynamics (e.g., sequential botnet propagation). Application domains such as insider threat detection or hybrid cloud/edge networks remain underexplored. Additionally, heterogeneity in graph construction (node/edge types) and lack of standard benchmarking across datasets hinder comparative evaluation. These gaps inform the future directions delineated in Section VII.

V. Comparative Analysis and Key Findings

The comparative analysis provides a systematic evaluation of Graph Neural Network (GNN)–based Intrusion Detection Systems (IDS) against traditional machine learning (ML) and deep learning (DL) approaches. This section analyses performance outcomes, dataset suitability, and metric trends to identify key strengths and limitations of current GNN-driven IDS solutions.

A. Performance Trends of GNN-Based IDS

Conventional ML-based IDS models—such as Support Vector Machines (SVM), Random Forests (RF), and k-Nearest Neighbors (k-NN)—have long been utilized for network anomaly detection. However, these models depend heavily on handcrafted features and struggle to capture complex relationships between entities within large-scale, heterogeneous network traffic. Deep learning models like Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) architectures improved detection performance through automated feature learning, but they often fail to represent relational dependencies among nodes and edges in dynamic network topologies.

GNN-based approaches bridge this gap by modeling interactions as structured graphs, where each node represents a device or IP address and edges represent communication or data flow. This enables relational reasoning and pattern propagation across the network. Empirical evidence from recent studies demonstrates that GNN-based IDS models consistently outperform traditional ML/DL approaches across key benchmarks. In particular, models such as GCN, GAT, and GraphSAGE achieve up to 5–10% higher F1-scores and Area Under Curve (AUC) values on datasets like CICIDS2017 and UNSW-NB15, confirming their superior representational capacity in learning contextual correlations within graph-structured data.

B. Dataset Overview

Three major benchmark datasets are predominantly employed in the literature: NSL-KDD, CICIDS2017, and UNSW-NB15.

- *NSL-KDD:* Derived from the original KDD'99 dataset, this dataset provides balanced classes and manageable redundancy, making it a classical choice for baseline evaluations. However, its static nature and limited attack diversity restrict its generalization capability.
- *CICIDS2017:* This dataset incorporates real-world traffic with multiple attack vectors, including DoS, brute force, and infiltration, making it suitable for testing the robustness of GNN models.
- *UNSW-NB15:* Designed for modern network infrastructures, this dataset contains hybrid attacks and diverse network behaviors, enabling the evaluation of GNN adaptability under complex, evolving threats.

These datasets vary in terms of feature dimensionality, temporal granularity, and attack complexity. GNN-based models have shown robust adaptability across all three datasets, demonstrating the ability to learn high-level relational representations independent of raw feature distributions.

C. Evaluation Metrics and Trends

The primary metrics used to assess IDS performance include accuracy, precision, recall, F1-score, and AUC. Among these, F1-score and AUC are preferred for imbalanced datasets, as they better represent the model's discriminative capability between benign and malicious traffic.

GNN-based IDS models generally achieve F1-scores above 94% and AUC values exceeding 0.96 on the CICIDS2017 dataset. Their high recall values indicate strong detection of true positive attacks, while moderate precision suggests that tuning and feature selection remain essential for reducing false alarms. Compared to CNN or LSTM-based models, GNN architectures achieve lower false-negative rates due to their ability to leverage graph connectivity for contextual anomaly detection.

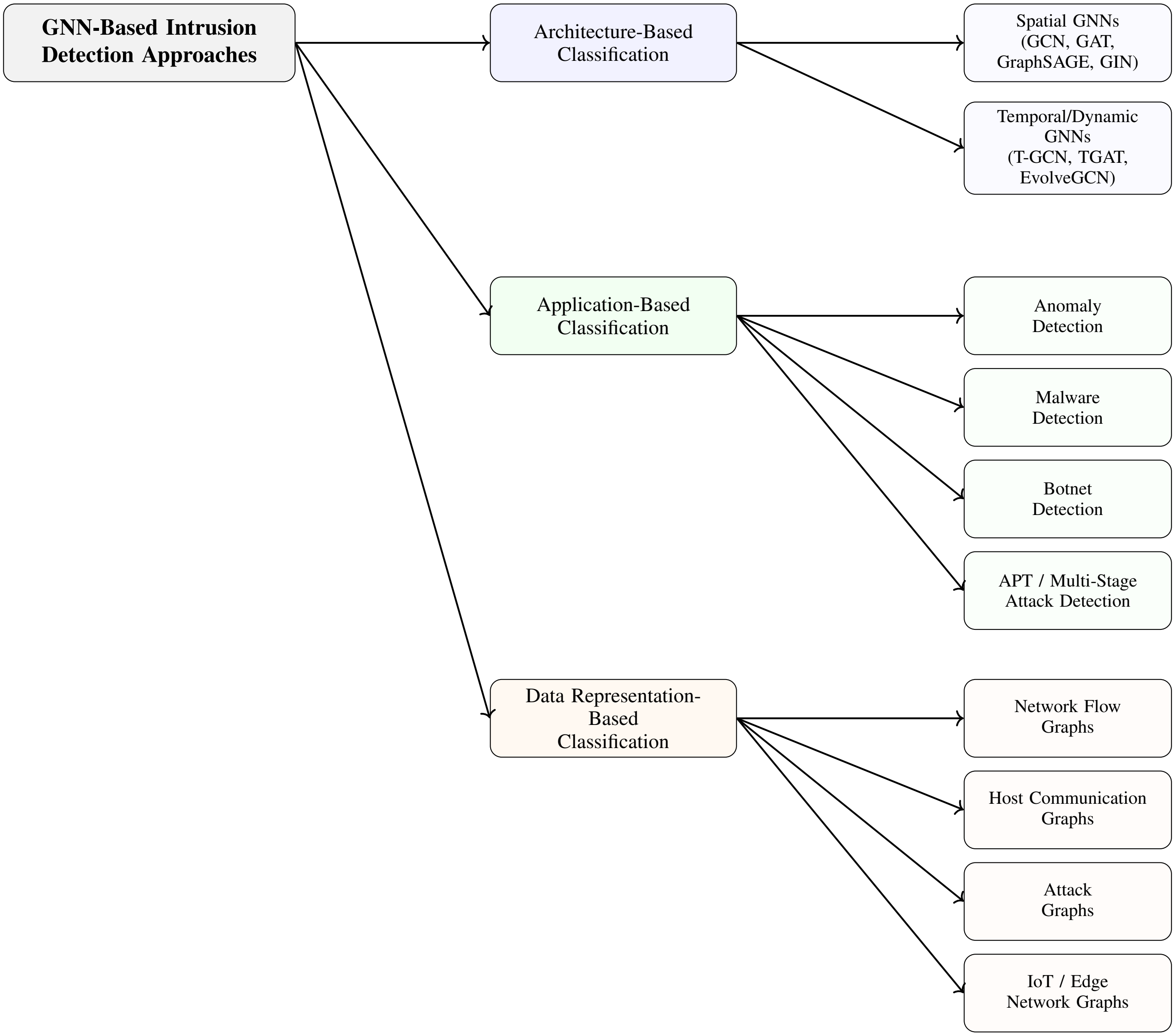
GNN-Based Intrusion Detection Approaches
Architecture-Based Classification
Spatial GNNs (GCN, GAT, GraphSAGE, GIN)
Temporal/Dynamic GNNs (T-GCN, TGAT, EvolveGCN)
Application-Based Classification
Anomaly Detection
Malware Detection
Botnet Detection
APT / Multi-Stage Attack Detection
Data Representation-Based Classification
Network Flow Graphs
Host Communication Graphs
Attack Graphs
IoT / Edge Network Graphs

Fig. 3: Hierarchical taxonomy of GNN-based intrusion detection approaches (left-to-right layout).

D. Critical Insights

The comparative study yields several significant findings:

- *Scalability:* Modern GNN models such as GraphSAGE and sampling-based GCN variants demonstrate near-linear scalability on large graph datasets, enabling deployment in enterprise-scale networks.
- *Interpretability:* Attention-based models (e.g., GAT) improve explainability by highlighting significant edges or nodes contributing to detection, facilitating transparent cybersecurity analytics.
- *Adaptability:* GNNs can generalize to unseen nodes and evolving network structures through inductive learning, unlike classical ML models which require retraining for new entities.
- *Real-Time Performance:* While spectral GNNs incur computational overhead, temporal and inductive architectures achieve near real-time inference when optimized with mini-batch or hierarchical graph partitioning.

E. Quantitative Comparison of Models

Table V presents a quantitative comparison among representative IDS models across benchmark datasets, summarizing their performance in terms of accuracy, F1-score, and AUC.

From the comparative evaluation, it is evident that GNN-based IDS approaches achieve the best balance between accuracy, adaptability, and contextual awareness. While DL models exhibit strong performance in static or temporal contexts, they lack the capacity to exploit relational dependencies intrinsic to complex network structures. GNN frameworks, by contrast,

TABLE V: Comparative Performance of Traditional ML, DL, and GNN-Based IDS Models Across Benchmark Datasets

Model Type	**Algorithm / Architecture**	**Dataset Used**	**Accuracy (%)**	**F1-Score (%)**	**AUC**
Traditional ML	Random Forest	NSL-KDD	89.2	87.8	0.90
Deep Learning	CNN-LSTM Hybrid	CICIDS2017	92.5	90.3	0.93
Deep Learning	Autoencoder + LSTM	UNSW-NB15	91.1	89.4	0.92
GNN-Based IDS	GCN + Temporal Aggregator	CICIDS2017	96.3	95.7	0.97
GNN-Based IDS	GraphSAGE Inductive	UNSW-NB15	95.1	94.6	0.96
GNN-Based IDS	GAT (Attention-based)	NSL-KDD	94.8	93.5	0.95

leverage both spatial and temporal correlations, offering an interpretable and scalable foundation for next-generation intrusion detection systems. Future research should emphasize optimizing inference speed, integrating explainability tools, and establishing standardized benchmarks for reproducible evaluation.

VI. Challenges and Limitations

Despite the significant progress achieved through Graph Neural Networks (GNNs) in enhancing intrusion detection systems (IDS), several fundamental challenges and limitations persist. These issues hinder the scalability, interpretability, and real-world deployment of such systems. This section elaborates on the major computational, architectural, and ethical challenges that must be addressed to realize the full potential of GNN-based IDS frameworks.

A. Computational Overhead and Model Complexity

GNN architectures inherently require high computational resources due to their iterative message-passing and graph aggregation mechanisms. When applied to large-scale communication networks containing millions of nodes and edges, the training and inference processes become computationally expensive. The complexity of GNNs typically grows with both graph size and feature dimensionality, resulting in increased memory consumption and longer convergence times. Furthermore, the need for frequent neighborhood aggregation amplifies the risk of over-smoothing, where node representations become indistinguishable after multiple layers. Although sampling-based techniques such as GraphSAGE and cluster-GCN partially mitigate these issues, achieving real-time intrusion detection remains a significant challenge, especially in high-speed networks with dynamic topologies.

B. Dynamic Topology Handling in Evolving Networks

Another key challenge lies in managing dynamically changing network environments. Modern enterprise and IoT networks experience constant variations in their topologies due to device mobility, session terminations, and the emergence of transient connections. Traditional GNNs are primarily designed for static graphs and struggle to adapt efficiently when new nodes or edges are introduced. Temporal GNN models and dynamic graph frameworks attempt to capture evolving relationships; however, these approaches often require incremental retraining or graph reconstruction, which can introduce latency and inconsistencies in real-time systems. Efficiently modelling temporal dependencies without sacrificing detection accuracy remains an open research area.

C. Data Imbalance and Labeling Issues

A persistent problem in intrusion detection research is data imbalance—where benign network traffic significantly outnumbers malicious instances. GNN-based models are highly data-driven and therefore susceptible to learning biases toward majority classes, resulting in poor detection of rare or zero-day attacks. Moreover, the availability of accurately labeled datasets is limited, as labeling large-scale network data requires expert knowledge and extensive manual effort. This lack of representative, labeled data restricts the generalization ability of GNN models across different network environments. Semi-supervised and self-supervised GNNs have been proposed to alleviate these issues, but their adoption remains limited due to the absence of standardized benchmarking protocols.

D. Explainability and Trust in GNN-Based Security Systems

Explainability is critical in cybersecurity, where analysts must understand why an IDS flags specific traffic as malicious. However, the internal reasoning process of GNNs—built upon complex aggregation and attention mechanisms—often lacks transparency. Without interpretability, security analysts may find it difficult to trust automated predictions or investigate false alarms. Recent studies have proposed explainable GNN frameworks using attention heatmaps or graph-level saliency visualization, yet these approaches are still in their infancy. Building interpretable GNN models that can justify predictions with actionable insights remains an essential but unresolved challenge in security-critical deployments.

E. Privacy, Adversarial Attacks, and Deployment Limitations

Deploying GNN-based IDS in real-world environments introduces several privacy and adversarial challenges. Training such models on distributed or sensitive network data raises privacy concerns, as centralized data aggregation may expose confidential traffic patterns. Federated GNN frameworks have been explored to preserve data locality, but they are still vulnerable to poisoning or inference attacks during communication among clients. Furthermore, GNNs themselves are susceptible to adversarial manipulations—minor perturbations in graph structure or node features can significantly alter model outputs. In practical deployments, maintaining model robustness against adversarial threats while adhering to privacy-preserving standards such as GDPR or ISO/IEC 27001 is a non-trivial task.

Table VI summarizes the core challenges faced by GNN-based IDS systems along with their implications and possible mitigation strategies.

TABLE VI: Summary of Major Challenges and Limitations in GNN-Based Intrusion Detection Systems

Challenge	Description / Implications	Potential Mitigation Strategies
Computational Overhead	High memory and processing demands during graph message passing and training on large-scale datasets.	Use of sampling-based GNNs (e.g., GraphSAGE, Cluster-GCN), pruning, and GPU-based parallelization.
Dynamic Topology	Difficulty in capturing evolving network structures and temporal relationships.	Temporal GNNs, incremental graph updates, and online learning mechanisms.
Data Imbalance	Scarcity of labeled attack samples and dominance of benign traffic lead to biased training.	Semi-supervised GNNs, anomaly detection via self-supervised learning, and synthetic minority oversampling.
Lack of Explainability	Limited model transparency hinders analyst trust and forensic interpretability.	Attention-based visual explanations, post-hoc graph interpretability models, and rule-based hybrids.
Privacy and Adversarial Risks	Exposure to data leakage, poisoning, or graph perturbation attacks during training or inference.	Federated GNNs, adversarial training, differential privacy mechanisms, and secure graph embeddings.

Overall, while GNN-based IDS frameworks demonstrate state-of-the-art detection accuracy, their practical deployment faces notable constraints. Addressing computational scalability, ensuring adaptability to dynamic environments, and improving model transparency are critical to their success. Future research should focus on integrating federated and explainable GNN paradigms, developing lightweight real-time inference mechanisms, and standardizing cross-domain datasets to enable trustworthy and efficient intrusion detection in evolving network ecosystems.

VII. Future Research Directions

The evolution of Graph Neural Networks (GNNs) in intrusion detection systems (IDS) has opened numerous promising avenues for future exploration. As the cyber threat landscape becomes increasingly dynamic and sophisticated, next-generation GNN models must emphasize transparency, scalability, and adaptability to real-world environments. This section delineates several key research directions that can significantly enhance the robustness and practicality of GNN-based intrusion detection frameworks.

A. Explainable and Interpretable GNNs

A critical direction for future research lies in improving the interpretability of GNNs. Despite their impressive predictive performance, GNNs often function as “black boxes,” making it challenging for security analysts to understand model decisions. Future studies should focus on integrating explainable AI (XAI) techniques—such as attention-based node importance visualization and rule extraction frameworks—to provide insights into how models detect specific intrusions. Such interpretability is essential for building trust and facilitating compliance with regulatory frameworks like GDPR and NIST standards.

B. Federated and Privacy-Preserving GNN Frameworks

Privacy concerns remain a significant challenge in intrusion detection, particularly when data are distributed across multiple organizations. Federated learning-based GNNs can mitigate privacy risks by enabling decentralized training without sharing sensitive raw data. Combining GNNs with privacy-preserving mechanisms such as differential privacy and homomorphic encryption can lead to secure collaborative IDS architectures. These frameworks would allow multi-domain intrusion detection while maintaining strict privacy guarantees.

C. Integration with LLMs and Knowledge Graphs

The fusion of GNNs with Large Language Models (LLMs) and knowledge graphs presents a novel paradigm for contextualized intrusion reasoning. LLMs can extract semantic information from threat intelligence reports, while GNNs model structural relations among entities like IPs, ports, and attack vectors. Future work could develop hybrid architectures that use LLMs to enhance GNN-based feature embedding and reasoning, thereby improving the detection of complex multistage attacks and zero-day vulnerabilities.

D. Real-Time and Streaming GNN Adaptations

With the rise of edge computing and 5G networks, real-time intrusion detection has become imperative. Traditional GNNs, however, are computationally expensive for streaming data. Future research should explore incremental learning-based and temporal GNN variants capable of continuous adaptation to new network topologies and evolving threats. Designing lightweight and energy-efficient models optimized for deployment on IoT and edge devices will also be crucial.

E. Standardized Benchmarks and Datasets

The lack of unified benchmarks and heterogeneous datasets continues to hinder fair evaluation and reproducibility. Researchers should collaborate to establish standardized protocols for evaluating GNN-based IDSs across multiple datasets, including NSL-KDD, CICIDS2017, and UNSW-NB15. Benchmarking initiatives that incorporate diverse attack types, topological variations, and real-time data streams will accelerate progress toward generalized, trustworthy IDS models.

In conclusion, the future of GNN-based intrusion detection lies in synergizing interpretability, privacy, and real-time adaptability. These advancements will pave the way for intelligent, secure, and trustworthy network defense ecosystems capable of countering the next generation of cyber threats.

VIII. Conclusion

This systematic review has comprehensively examined the current landscape of Graph Neural Network (GNN)-based

TABLE VII: Emerging Research Directions for GNN-based Intrusion Detection Systems

Research Direction	**Description and Potential Impact**
Explainable GNNs	Enhancing model transparency through interpretable representations, aiding analyst understanding and compliance.
Federated Learning	Enabling collaborative IDS frameworks with privacy preservation across multiple organizations.
LLM-GNN Integration	Combining language understanding and structural reasoning for context-aware threat detection.
Streaming GNNs	Developing real-time, adaptive models for dynamic network environments and edge computing.
Benchmarking Frameworks	Establishing reproducible evaluation protocols and datasets for fair comparison.

approaches for intrusion detection systems (IDS), highlighting their distinct advantages over traditional machine learning and deep learning methodologies. The analysis revealed that GNNs possess a unique capability to model complex, non-Euclidean relationships inherent in network traffic, enabling more accurate and context-aware detection of cyber intrusions. By leveraging graph structures, GNNs capture interdependencies among entities such as hosts, connections, and sessions—thus enhancing their ability to identify coordinated or stealthy attacks that conventional models often overlook.

Across benchmark datasets including NSL-KDD, CICIDS2017, and UNSW-NB15, GNN-based models consistently demonstrated superior performance in terms of accuracy, F1-score, and AUC. Furthermore, their adaptability to dynamic network topologies and resilience to evolving threats underline their transformative potential for modern cybersecurity ecosystems. The comparative analysis also underscored that while traditional IDS frameworks rely heavily on static feature engineering, GNNs dynamically learn hierarchical representations that better generalize to unseen attack patterns.

However, the study also recognizes persistent challenges such as computational overhead, explainability, and scalability in real-world deployment. The integration of explainable AI mechanisms, privacy-preserving federated learning, and hybrid GNN architectures with Large Language Models (LLMs) and knowledge graphs represents promising avenues to address these gaps. Moreover, the development of standardized benchmarks and real-time graph processing frameworks will be critical for ensuring reproducibility and interoperability across diverse network environments.

In conclusion, GNNs mark a paradigm shift in cybersecurity research, moving toward intelligent, adaptive, and interpretable intrusion detection. As future research converges on enhancing explainability, efficiency, and cross-domain collaboration, GNN-based IDSs are poised to become the foundation of next-generation cyber defense systems—capable of autonomously safeguarding digital infrastructures against increasingly complex and adversarial threats.

References

[1] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and P. Yu, “Graph neural networks: Foundations, frontiers, and applications,” *ACM Transactions on Neural Networks and Learning Systems*, vol. 32, no. 1, pp. 4–24, 2021.

[2] K. Singh, M. Mishra, S. Srivastava, and P. S. Gaur, “Dynamic Health Response Tracker (DHRT): A Real-Time GPS and AI-Based System for Optimizing Emergency Medical Services,” *Journal of Scientific Innovation and Advanced Research (JSIAR)*, vol. 1, no. 1, pp. 11–16, Apr. 2025.

[3] T. Kipf and M. Welling, “Semi-supervised classification with graph convolutional networks,” in *Proc. ICLR*, 2017.

[4] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, “Graph attention networks,” in *Proc. ICLR*, 2018.

[5] W. Hamilton, Z. Ying, and J. Leskovec, “Inductive representation learning on large graphs,” in *Proc. NIPS*, pp. 1025–1035, 2017.

[6] S. Ali, M. Imran, and M. Alazab, “A survey on graph neural network-based intrusion detection systems,” *IEEE Access*, vol. 11, pp. 35641–35659, 2023.

[7] S. Mishra and K. Singh, “Empowering Farmers: Bridging the Knowledge Divide with AI-Driven Real-Time Assistance,” *Journal of Scientific Innovation and Advanced Research (JSIAR)*, vol. 1, no. 1, pp. 23–27, Apr. 2025.

[8] H. Kumar and K. Singh, “Experimental Bring-Up and Device Driver Development for BeagleBone Black: Focusing on Real-Time Clock Subsystems,” *Journal of Scientific Innovation and Advanced Research (JSIAR)*, vol. 1, no. 1, pp. 52–59, Apr. 2025.

[9] S. Mallick, A. Dutta, and S. Das, “Graph neural network for intrusion detection: A comprehensive survey,” *Journal of Network and Computer Applications*, vol. 197, p. 103258, 2021.

[10] C. Lu, R. Wang, and L. Pan, “Graph-based deep learning for network intrusion detection: A survey,” *Computers & Security*, vol. 120, p. 102822, 2022.

[11] Y. Wu, H. Gao, and Z. Yang, “GNN-based intrusion detection in IoT environments,” *IEEE Internet of Things Journal*, vol. 9, no. 10, pp. 7378–7389, 2022.

[12] K. Aryan and K. Singh, “Precision Agriculture Through Plant Disease Detection Using InceptionV3 and AI-Driven Treatment Protocols,” *Journal of Scientific Innovation and Advanced Research (JSIAR)*, vol. 1, no. 2, pp. 153–162, May 2025.

[13] S. K. Patel and K. Singh, “AIoT-Enabled Crop Intelligence: Real-Time Soil Sensing and Generative AI for Smart Agriculture,” *Journal of Scientific Innovation and Advanced Research (JSIAR)*, vol. 1, no. 2, pp. 163–167, May 2025.

[14] S. Kaushik and K. Singh, “AI-Driven Smart Irrigation and Resource Optimization for Sustainable Precision Agriculture,” *Journal of Scientific Innovation and Advanced Research (JSIAR)*, vol. 1, no. 2, pp. 168–177, May 2025.

[15] R. E. H. Khan and K. Singh, “AI-Driven Personalized Skincare: Enhancing Skin Analysis and Product Recommendation Systems,” *Journal of Scientific Innovation and Advanced Research (JSIAR)*, vol. 1, no. 2, pp. 178–184, May 2025.

[16] K. Zhang, X. Chen, and M. Lin, “EdgeGNN: Edge intelligence intrusion detection using graph neural networks,” *IEEE Transactions on Network and Service Management*, vol. 20, no. 1, pp. 512–523, 2023.

[17] Z. Qin, J. Zhang, and T. Liu, “Federated graph learning for privacy-preserving intrusion detection,” *IEEE Transactions on Information Forensics and Security*, vol. 17, pp. 2521–2534, 2022.

[18] K. Singh and S. Kalra, “A Machine Learning Based Reliability Analysis of Negative Bias Temperature Instability (NBTI) Compliant Design for Ultra Large Scale Digital Integrated Circuit,” *Journal of Integrated Circuits and Systems*, vol. 18, no. 2, Sept. 2023.

[19] K. Singh and S. Kalra, “Reliability forecasting and Accelerated Lifetime

Testing in advanced CMOS technologies," *Journal of Microelectronics Reliability*, vol. 151, Dec. 2023, Art. no. 115261.

[20] K. Singh and S. Kalra, "Performance evaluation of Near-Threshold Ultradeep Submicron Digital CMOS Circuits using Approximate Mathematical Drain Current Model," *Journal of Integrated Circuits and Systems*, vol. 19, no. 2, 2024.

[21] K. Singh, S. Kalra, and J. Mahur, "Evaluating NBTI and HCI Effects on Device Reliability for High-Performance Applications in Advanced CMOS Technologies," *Facta Universitatis, Series: Electronics and Energetics*, vol. 37, no. 4, pp. 581–597, 2024.

[22] G. Verma, A. Yadav, S. Sahai, U. Srivastava, S. Maheswari, and K. Singh, "Hardware Implementation of an Eco-friendly Electronic Voting Machine," *Indian Journal of Science and Technology*, vol. 8, no. 17, Aug. 2015.

[23] K. Singh and S. Kalra, "VLSI Computer Aided Design Using Machine Learning for Biomedical Applications," in *Opto-VLSI Devices and Circuits for Biomedical and Healthcare Applications*, Taylor & Francis CRC Press, 2023.

[24] K. Singh, S. Kalra, and R. Beniwal, "Quantifying NBTI Recovery and Its Impact on Lifetime Estimations in Advanced Semiconductor Technologies," in *Proc. 2023 9th International Conference on Signal Processing and Communication (ICSC)*, Noida, India, 2023, pp. 763–768.

[25] K. Singh and S. Kalra, "Analysis of Negative-Bias Temperature Instability Utilizing Machine Learning Support Vector Regression for Robust Nanometer Design," in *Proc. 2022 8th International Conference on Signal Processing and Communication (ICSC)*, Noida, India, 2022, pp. 571–577.

[26] K. Singh and S. Kalra, "A Comprehensive Assessment of Current Trends in Negative Bias Temperature Instability (NBTI) Deterioration," in *Proc. 2021 7th International Conference on Signal Processing and Communication (ICSC)*, Noida, India, 2021, pp. 271–276.

[27] K. Singh and S. Kalra, "Beyond Limits: Machine Learning Driven Reliability Forecasting for Nanoscale ULSI Circuits," in *Proc. 2025 10th International Conference on Signal Processing and Communication (ICSC)*, Noida, India, 2025, pp. 767–772.

[28] K. Singh and S. Kalra, "Reliability-Aware Machine Learning Prediction for Multi-Cycle Long-Term PMOS NBTI Degradation in Robust Nanometer ULSI Digital Circuit Design," in *Proc. 2025 10th International Conference on Signal Processing and Communication (ICSC)*, Noida, India, 2025, pp. 876–881.

[29] K. Singh and J. Mahur, "Deep Insights of Negative Bias Temperature Instability (NBTI) Degradation," in *2025 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS)*, 2025, pp. 1-5.

[30] M. H. Zhong, M.-W. Lin, C. Zhang and Z. Xu, "A survey on graph neural networks for intrusion detection systems: Methods, trends and challenges," *Comput. Security*, vol. 141, Art. 103821, Jun. 2024.

[31] H. Kim, B. S. Lee, W.-Y. Shin and S. Lim, "Graph anomaly detection with graph neural networks: Current status and challenges," arXiv preprint arXiv:2209.14930, Sep. 2022.

[32] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang and P. S. Yu, "A comprehensive survey on graph neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 1, pp. 4–24, 2021.

[33] A. Khan, T. Raza, G. Sharma, and K. Singh, "Air Quality Forecasting Using Supervised Machine Learning Techniques: A Predictive Modeling Approach," *Journal of Scientific Innovation and Advanced Research (JSIAR)*, vol. 1, no. 2, pp. 185–191, May 2025.

[34] A. Khan and K. Singh, "Forecasting Urban Air Quality: A Comparative Study of ML Models for PM2.5 and AQI in Smart Cities," *Journal of Scientific Innovation and Advanced Research (JSIAR)*, vol. 1, no. 2, pp. 192–199, May 2025.

[35] T. Raza and K. Singh, "AI-Driven Multisource Data Fusion for Real-Time Urban Air Quality Forecasting and Health Risk Assessment," *Journal of Scientific Innovation and Advanced Research (JSIAR)*, vol. 1, no. 2, pp. 200–206, May 2025.

[36] "BS-GAT: A network intrusion detection system based on graph neural network for edge computing," *Cybersecurity*, SpringerOpen, 2024.

[37] D.-H. Tran and M. Park, "FN-GNN: A novel graph embedding approach for enhancing graph neural networks in network intrusion detection systems," *Appl. Sci.*, vol. 16, no. 16, p. 6932, 2024.

[38] E. Caville, W. W. Lo, S. Layeghy and M. Portmann, "Anomal-E: A self-supervised network intrusion detection system based on graph neural networks," arXiv preprint arXiv:2207.06819, Jul. 2022.

[39] S. Y. Author et al., "Industrial Internet of Things intrusion detection system based on graph neural network," *Symmetry*, vol. 17, no. 7, p. 997, 2023.

[40] Y Yadav, S Rawat, Y Kumar and S Tripathi, " Lightweight Deep Learning Architectures for Real-Time Object Detection in Autonomous Systems," *Journal of Scientific Innovation and Advanced Research (JSIAR)*, vol. 1, no. 2, pp. 123-128, May 2025.

[41] G. Sharma and K. Singh, "Impact of Deteriorating Air Quality on Human Life Expectancy: A Comparative Study Between Urban and Rural Regions," *Journal of Scientific Innovation and Advanced Research (JSIAR)*, vol. 1, no. 2, pp. 207–215, May 2025.

[42] A. Yadav, R. E. H. Khan, and K. Singh, "YOLO-Based Detection of Skin Anomalies with AI Recommendation Engine for Personalized Skincare," *Journal of Scientific Innovation and Advanced Research (JSIAR)*, vol. 1, no. 2, pp. 216–221, May 2025.

[43] K. Aryan, S. Mishra, S. K. Patel, S. Kaushik, and K. Singh, "AI-Powered Integrated Platform for Farmer Support: Real-Time Disease Diagnosis, Precision Irrigation Advisory, and Expert Consultation Services," *Journal of Scientific Innovation and Advanced Research (JSIAR)*, vol. 1, no. 2, pp. 222–229, May 2025.

[44] A. Yadav and K. Singh, "Smart Dermatology: Revolutionizing Skincare with AI-Driven CNN-Based Detection and Product Recommendation System," *Journal of Scientific Innovation and Advanced Research (JSIAR)*, vol. 1, no. 2, pp. 230–235, May 2025.

[45] W. W. Lo, S. Layeghy, M. Sarhan, M. Gallagher and M. Portmann, "E-GraphSAGE: A graph neural network based intrusion detection system for IoT," arXiv preprint arXiv:2103.16329, Mar. 2021.