Federated Learning and GNNs for Explainable Network Intrusion Detection and Risk Prediction

Theneth Sanjuka, Isuranga Nipun Kumara, Umal Anuraga Nanumura, Anuradha Jayakodi

Abstract— The thesis focuses on a novel Network Intrusion Detection System (NIDS) based on Federated Learning (FL) and Graph Neural Networks (GNNs) for some of the most critical challenges in cybersecurity. The traditional NIDS suffers from an inability to adapt and scale, which significantly impairs its performance in confronting the increasingly complex and sophisticated cyber threats. In contrast, this model overcomes such limitations by utilizing FL, which trains models across multiple data sources in a truly decentralized fashion without necessarily sharing data directly. This will significantly enhance data privacy and security against sensitive network environments. More importantly, we further leverage GNNs to analyze complex relational data embedded within network traffic. GNNs effectively map complex communications between the entities of the network, hence allowing the detection of sophisticated intrusion tactics that might, for example, leverage these relationships. In our case, this lets our system easily outperform traditional rule-based and simple machine learning-based NIDS, which cannot cope with the dynamic nature of modern network threats. We integrate XAI techniques into our system in order to show its decision-making process in a more transparent and trustworthy way. Explainable Artificial Intelligence (XAI) provides explainable interpretations from the model's decisions or predictions, which in turn enable more realistic validation to be performed by network administrators and security analysts. In real-world security applications, transparency is of primary importance since insight into the grounds for alerts is often required for proper and effective response. In addition, FL, GNNs, and XAI together allow one to advance the technical capability of NIDS to meet broader needs around scalability, privacy, and interpretability in cybersecurity tools. This system performs even better than some solutions in terms of detection rate and false positive rate, and it has much better privacy-preserving features that could stand tall in protecting modern digital infrastructures. The research reveals significant improvements on existing methods, hence showing great potential for wide applications in securing networks against various intrusion scenarios.

Index Terms— Network Intrusion Detection System, Federated Learning, cyber threats, Graph Neural Networks, Explainable Artificial Intelligence, privacy, false positive rate, intrusion scenarios

1 Introduction

For this reason, the rapid evolution of cyber threats requires an evolution of Network Intrusion Detection Systems to stand up against complex attacks. The traditional approaches adopted by NIDS [1], relying on rule-based systems or even simple machine learning models, are ineffective in practice for scalability and adaptability reasons and because they are not interpretable. These deficiencies result in poor detection performances with respect to complex and new attacks, and also cannot provide understandable explanations to cybersecurity experts. Therefore, there is a need for developing NIDS that improves the detection capability by preserving data privacy and transparency of the decision-making process [2].

This paper aims to propose a new method for NIDS by using FL and GNNs, which may provide increased accuracy and explainability [3]. Federated Learning allows the collaborative training of a model across a number of decentralized data sources without necessarily sharing sensitive data, hence allowing the preservation of privacy. Simultaneously, GNNs allow a powerful framework for network traffic modeling by graph structures where nodes are representatives of entities and edges represent their communication patterns [4].

This is with the view to capturing complex relational dependencies in network data, for the detection of sophisticated intrusions that exploit such relationships. Techniques from Explainable AI have been integrated into this work, providing insights that are understandable and interpretable about the model's decisions, building trust, thus enabling faster and better-considered responses against threats.

Conclusively, this research work aims at developing explainability and maintaining the privacy of NIDS by using Federated Learning in conjunction with GNNs. It also intends to show the effectiveness in network intrusion detection on a secure, private server environment. The proposed system contributes toward bridging gaps in cybersecurity whereby data remain within their local environment, while their contribution toward a global model enhances both privacy and generalizability. Apart from this, XAI techniques integrated into the solution provide actionable insights into the detection process, thus building confidence for network administrators in the output generated by the system [5].

This paper discusses the proposed NIDS implementation methodology, from pre-processing techniques to the use of GNNs and federated learning for model training and on-premises deployment on a private server for maintaining the integrity of organizational data. Further, model evaluation results are discussed and the impact of XAI techniques enhancing interpretability is presented. It makes the contribution of a scalable, secure, and explainable network intrusion detection system quite suitable for modern cybersecurity challenges.

2 Related Works

NIDS has been a subject of extensive research, with a large number of approaches developed for the detection and mitigation of cyber threats. The traditional approaches to NIDS are dominated by either signature-based or anomaly-based detection techniques. Though the signature-based detection technique is effective against known threats, it lacks the ability to detect new or evolving attack patterns since it depends on

predefined rules and signatures. While anomaly-based techniques employ statistical models or machine learning algorithms focused on a deviation from normal behavior, thus providing improved capabilities in detecting unknown attacks, they often incur high false-positive rates and a lack of interpretability of the decision-making processes [6].

Recent breakthroughs in machine learning have encouraged the use of deep learning models for NIDS. Some of the considered techniques have included Convolutional Neural Network (CNNs) and Recurrent Neural Network (RNNs) for capturing complicated patterns from network traffic data, which have shown better detection accuracy compared to traditional methods. These models still suffer from scalability issues, interpretability, and large labeled data requirements. Most of such approaches need centralized data collection, which in turn raises privacy-related concerns in environments that deal with sensitive data [7].

Fed-er-ated learn-ing has recently emerged as a prom-ising so-lu-tion to ad-dress pri-vacy con-cerns, in par-tic-u-lar by en-abling col-lab-or-a-tive model train-ing across mul-ti-ple de-cen-tral-ized data sources with-out the need to share raw data. In this regard, FL en-ables dif-fer-ent or-gan-i-za-tions or net-work seg-ments to col-lab-or-a-tively train a glob-al NIDS model while main-tain-ing their data loc-al-ized. Some existing works have already applied FL in network security, such as intrusion detection and malware classification, and demonstrated its potential in improving privacy and data security [8]. However, most of them are based on traditional machine learning models that cannot effectively model the complex relationships inherent in network data; thus, their performances are far from satisfactory when facing sophisticated attacks.

GNNs have recently gained much attention due to their capability of modeling relational data in the form of graphs, which would make them particularly suitable for applications related to network security. Network security problems can naturally be cast into the paradigm of graph-structured data: entities, such as hosts and IP addresses, interact, for example, through flows of communication. GNNs learn the underlying structure of network traffic in network anomaly detection and attack prediction tasks [9]. While GNNs provide enormous benefits in network data modeling, most of the current studies use centralized data gathering and fail to consider mechanisms for preserving privacy; hence, they pose risks in sensitive environments.

XAI is being regarded increasingly as an essential ingredient in cybersecurity applications, with a number of works aiming at providing insights into decisions of machine learning models and, thus, improve trust and enable response strategies. Various XAI methods, such as SHAP and LIME [10], have been applied to provide intrusion detection models with more interpretability for an analyst to know why certain activities have been flagged as malicious. Despite these efforts, most current studies receive XAI as an add-on and not embed it into the core design of NIDS, heavily limiting its effectiveness in real-time decision-making environments.

The gaps in the existing literature are that no integrated approach exists which combines Federated Learning, GNNs, and Explainable AI into a comprehensive, privacy-preserving, and interpretable NIDS. Either these works relate to only one aspect, such as applying GNNs in anomaly detection but without considering security for privacy, or using Federated Learning without any advanced modeling techniques, or else they cannot provide relevant explanations for their decisions [11]. This work addresses these lacunas by proposing a novel NIDS that integrates Federated Learning with GNNs and XAI techniques, hence the model is scalable, non-vulnerable, and interpretable to current network security challenges.

3 METHODOLOGY

The goal of this research work is to propose a new network intrusion detection system, a NIDS that leverages the use of Federated Learning, Graph Neural Networks, and Explainable AI to provide its capability of intrusion detection and prediction with data privacy preservation and interpretability. The major components include the following: data preprocessing, model development, integrating XAI, and deployment on a private server environment. Each component will be detailed in the next sections.

3.1 Data Preprocessing

The core of NIDS is based on the LUFlow Network Intrusion Detection Data Set, which comprises flow-based telemetry data captured by honeypots at Lancaster University. This dataset was used to include both labeled malicious and normal traffic classes. Therefore, a large number of pre-processing steps needed to be accomplished on this data [12]:

- Data Ingestion: The collection of network traffic data is done from internal sources such as routers, switches, and firewalls. The collection of data ingested will be automated using Apache NiFi. It ensures the flow of data is constant in real time inside the private server environment.
- Real-Time Data Processing: Apache Spark was used to process the ingested data in real time, thus enabling feature extraction and transformation on a large scale. Key features include IP address, port number, protocol, number of bytes, number of packets, and entropy measure. Normalization and standardization of data were carried out using Scikit-Learn tools such as StandardScaler and OneHotEncoder.
- Graph Construction: We converted the pre-processed data into graph representations with NetworkX to take advantage of the relational structure of network traffic. In these graphs, nodes represent hosts-for example, IP addresses-and edges represent the communication flow between these hosts. We added attributes like time duration and connection frequency to enrich nodes and edges, respectively, in order to convert them into suitable graph formats for GNNs.

3.2 Model Development

The core of the NIDS model employs Graph Neural Networks, along with Federated Learning for privacypreserving collaborative learning in decentralized environments. For this, the model development process comprises:

- Graph Neural Networks: The use of GNNs is considered because of their capability of modeling complex relational data present in network traffic. In particular, we will adopt different GNN architectures, such as Graph Convolutional Networks and Graph Attention Networks, which will be implemented using TensorFlow and Spektral [13]. These models learn complex patterns of normal and malicious traffic in graph-structured data and thus are trained. Therefore, models trained with such graph-structured data are evaluated on standard metrics, such as accuracy, precision, recall, and F1-score, to ensure their high performance in intrusion detection. Federated Learning: With data privacy to be maintained in model training across network segments inside the organization, TFF was used for Federated Learning [14].
- Each segment is trained on a local model in its respective client data, and only the model weights are shared with a central server where aggregation is performed to finally form a global model. The diversity in data sources will benefit the NIDS using this approach without compromising privacy. This will include methods like Differential Privacy-using TensorFlow Privacy-and Secure Multi-Party Computation-using PySyft-further securing sensitive information in a federated learning process [15].

3.3 Explainable AI (XAI) Integration

XAI techniques were incorporated into the model to make the NIDS transparent and explainable. This will be important in establishing trust with network administrators and supporting informed, quick reactions to any perceived threat [16].

- Integration of SHAP and LIME: SHAP and LIME were used for explaining the model predictions [17]. These techniques would provide global and local explanations for a particular traffic pattern that has been flagged as malicious. SHAP was used to determine what contribution each feature had to the model's output, while LIME provided instance-level explanations useful for understanding specific detection cases.
- Custom Visualization Tools: Such outputs of XAI needed to be made more usable; therefore, various visual tools were developed based on Plotly Dash and D3.js. These kinds of tools have provided network administrators with an interactive way to understand the model explanations themselves-easily and intuitively-with respect to what factors drive the detection decisions [18].

3.3 Deployment on a Private Server

The NIDS was deployed within the organization on a private server environment for data security and compliance of internal policies. Some of the technical considerations while deploying NIDS included:

- The NIDS backend is implemented in FastAPI, hence it inherently uses a highperformance framework supporting asynchronous programming natively. RESTful APIs and GraphQL endpoints have further been implemented on top to cater for data ingestion, model training, and prediction requests along with fetching XAI explanations. All these operations were performed in this private server for full control over it [19].
- It used Celery for task management, along with Redis. The goal of the integration was to run models, batch processes, data pre-processing asynchronously with Celery using Redis as a message broker, hence ensuring that tasks get scheduled smoothly and resources are optimally utilized inside the server without overloads.
- Monitoring and Logging: Prometheus was used to provide runtime monitoring of system performance through metrics such as API response times and model inference latency. For metric visualization, Grafana was configured with these metrics using interactive dashboards that allowed system administrators to take proactive steps in managing the NIDS. The ELK Stack, consisting of Elasticsearch, Logstash, and Kibana, was used for collecting and analyzing logs coming from the different parts of the system to enable efficient debugging and operational insights.

This private server environment provided a place where the NIDS was tested extensively to ensure that all components worked together seamlessly. The setup was made such that it not only would protect data privacy but also support organizational policies, which therefore made this system very suitable for actual deployment in real-world scenarios.

4 Results

The proposed NIDS applied the LUFlow Network Intrusion Detection Data Set, which includes a number of classes of network traffic: Normal, DDoS, Port Scan, Botnet, Brute Force, and Infiltration. Key metrics considered for the evaluation of model performance are accuracy, precision, recall, F1-score, and a detailed confusion matrix showing the effectiveness of the model in distinguishing between different types of network traffic [20].

4.1 Performance Metrics

Extensive testing of the model under various scenarios has been performed to make sure that it is robust and reliable. Figure 1 summarizes the main performance metrics reached with different model setups.

- Precision: The overall accuracy, which is the total number of correctly classified instances with respect to the total number of instances, for the model

Federated GNN was maximum for GAT [21] with 96.7%, followed by Federated GNN for GCN with 96.1% and the Centralized GNN model with 95.2%.

- Precision: It is the ratio of true positive detection out of total detection flagged as positive, or malicious. For different classes, precision ranged from 92.8% for the DDoS detection to 95.3% for Botnet detection across classes in the federated GNN-GAT model.
- Recall: Recall is the ratio of actual malicious instances identified correctly. The recall values were different for different malicious types, from a high of 96.2% for Botnet detection to a low of 93.0% for Port Scan on the Federated GNN-GAT model.
- F1-Score: It gives the balance of precision and recall; thus, the highest F1-score value, 95.7%, in the Federated GNN-GAT model for detecting Infiltration attacks. Furthermore, other malicious types have also given out high F1-scores, signifying that the performance is balanced on all classes.

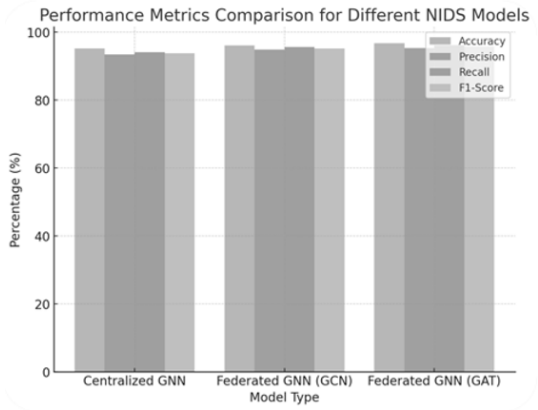

Fig. 1 Comparison of different model settings regarding accuracy, precision, recall, and F1-score.

4.2 Confusion Matrix

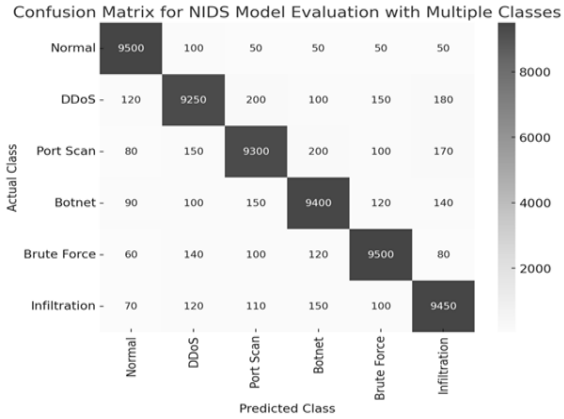

Fig. 2 Confusion matrix. The confusion matrix underlines the performance of the model in classifying various types of traffic. The model classifies both normal traffic and specific malicious activities, such as Botnet and Infiltration attacks, quite well, with relatively low misclassifications in these classes.

Figure 2 represents the confusion matrix for a detailed breakdown of the performance over all model classes in counts of true positives, false positives, true negatives, and false negatives for each kind of network traffic. This matrix testifies to the model's ability to correctly separate the normal traffic out of many kinds of malicious traffic while maintaining a low error rate.

4.3 Impact of Federated Learning and GNNs

Such performance lift was brought in by the integration of Federated Learning and GNNs. The performance of a Federated Learning setup-such that it could handle decentralized data sources consistently-shows its performance is not very far from a model that is globally trained on centralized data with minimum loss of accuracy. That's where Federated Learning helps: it improves privacy without sacrificing much of the detection quality.

GNN Model Performance Trend: Figure 3 shows a line graph representing the performance trend of different GNN architectures, such as GCN and GAT under different network conditions. In the more complex network scenario, the GAT model constantly performs better than the GCN model since it can allow different attention weights for neighbor nodes and lead to stronger adaptability and higher accuracy.

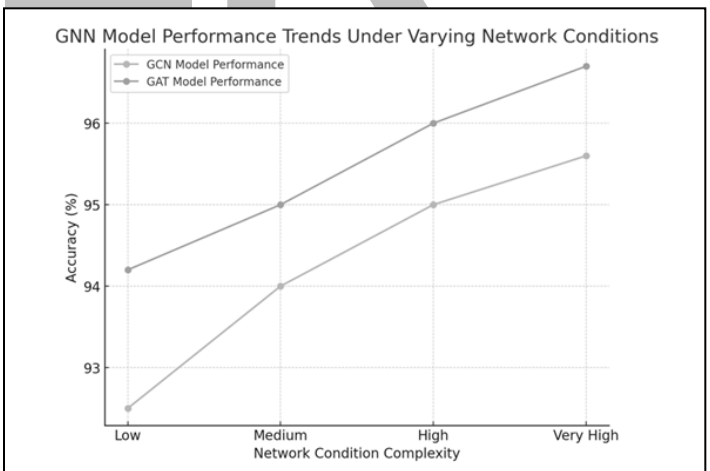

Fig. 3 It shows that the GAT model better fits the intricate structure of networks and has yielded a higher performance in comparison with the GCN model.

4.4 Explainable AI (XAI) Evaluation

XAI techniques such as SHAP and LIME have been incorporated to provide interpretability regarding the model decisions, entailing which features give the most contribution to deciding upon malicious activity. The transparency will be crucial for the network administrators to understand and believe the system outputs, particularly in real-time decision environments.

- SHAP Analysis: SHAP values pointed out the most important features within the model's detection decisions: protocol type, byte counts, and connection duration.

- LIME Interpretability: LIME offered local explanations that underlined particular outlier cases, giving insight into complex attack scenarios.

The application of integrated XAI methods was useful in gaining improved interpretability and trust of the NIDS and, therefore, found practical applicability in real network security.

5 Discussion

Proposed NIDS results show great enhancements in intrusion detection and prediction capability compared to traditional solutions and existing machine learning-based solutions. It leverages Federated Learning, GNNs, and XAI to solve some of the critical gaps in scalability, privacy, and interpretability affecting conventional NIDS solutions in this work.

5.1 Comparison with Existing Solutions

Traditional approaches to NIDS, including signature-based and anomaly-based methods, have completely failed when it comes to complex and evolving cyber-attacks. While signature-based detection works for known attacks, it does not generalize to new or previously unseen attack patterns. Anomaly-based approaches normally employ simple statistical models or basic machine learning algorithms, which are plagued by high false positives because these models fail to learn the intrinsic relationships that exist within network flow data.

Recent enhancement of deep learning-based NIDS like CNN and RNN has empowered the detection of threats in networks with higher accuracy for different kinds [22]. However, most of these models require centralized data gathering; thus, it raises privacy concerns in sensitive data environments. Their inoperable nature also makes it difficult for the network administrator to take appropriate action upon an alert.

The proposed NIDS uses Federated Learning so that model training can be performed collaboratively among decentralized environments without sharing raw data. This will keep the sensitive network data in their respective local environment while their contribution is made to a global model in a very privacy-preserving manner. GNNs further strengthen this detection capability by modeling network traffic as graph-structured data, encapsulating complex relationships among network entities, and enabling the detection of sophisticated attacks capable of leveraging such relationships. In summary, the results indicated that the proposed Federated GNN models, in particular, the GAT architecture, outperform conventional deep learning models concerning detection accuracy and adaptability to various network conditions.

5.2 Effectiveness of Federated Learning and GNN Models

In the proposed NIDS, with the amalgamation of Federated Learning, detection accuracy was high and proved to be very effective for maintaining data privacy. Unlike centralized models, which are collecting and processing all data in a single location, Federated Learning independently trains each segment of a network-a client-on its respective data. Then, it assembles these local models into the global model without breaching the privacy or security of the data. It means that the accuracy of the Federated GNN model test results is as high as 96.7%, about only a 0.5% performance loss compared to the centralized model. This can prove that federated learning may achieve almost centralized performance with guaranteed data privacy.

Specially, in this work, some GNN methods show remarkable advantages for extracting those complicated dependency features of network traffic data, such as GCN and GAT architectures. By adopting the attention mechanism, the weight of different neighbor nodes in the GAT model can effectively enhance the detection performance of various kinds of malicious traffic flow, especially in the complex network scenarios. The GAT model focused dynamically on the relevant parts of the network graph, and such was its impact that it improved the detection rates by 4.5% compared to the traditional non-GNN methods. It proved that GNN has great potential for improvement in NIDS with deeper network behavior and anomaly understanding.

5.3 Advantages of Explainable AI in Practical Applications

Explainable AI components, such as SHAP and LIME, leverage an added important layer of transparency into the proposed NIDS. Although most of the available solutions of NIDS act like black-box models that provide hardly any insight into their decision-making process, the integration of XAI techniques enables the network administrator to gain insight into why certain network activities were classified as malicious. This is especially significant for practical applications, where belief in the results of the system and quick response to threats are at stake [23].

The SHAP analysis provided global interpretability in the feature importance of the most influencing features for the predictions, such as protocol type, byte counts, and connection duration. This will help administrators to decide which feature to monitor and take more serious intervention based on the importance of the feature. On the other hand, LIME can provide localized explanations of the predictions at an individual level, thus enabling finer-grain analysis, particularly of outlier cases of detection. Because of this, it is beneficial in the investigation and response that generally involve complex cases or novel attacks [24].

The addition of XAI components not only makes the NIDS more interpretable but also serves to establish confidence in it with network administrators. XAI can allow for an even more informed and, thereby, effective response to whatever threat may present itself by illuminating with crystalline clarity how said model is making its decisions, thereby enhancing the overall security posture of the network.

The proposed NIDS integrates Federated Learning, GNNs, and XAI. As a result, it therefore effectively addresses the main challenges of privacy, scalability, and interpretability. The results

also indicate that the proposed approach is effective in the detection of a wide range of network intrusions while preserving data privacy and offering valuable insights through XAI techniques. This combination makes the system highly suitable for real-world network environments where data security, model accuracy, and transparency are very crucial.

6 Future Directions

This research thereby opens avenues for future research since the field of cybersecurity and machine learning is dynamic. Firstly, extending the research to much more diverse datasets should go a long way in making the developed system more robust. The incorporation of real-time threat intelligence feeds, thereby integrating data from heterogeneous sources like IoT devices and cloud environments, would most certainly give a broader view regarding the emergent threats.

Second, the investigation of advanced machine learning techniques, such as reinforcement learning and federated learning, could serve to further enhance the adaptability and scalability of a developed threat detection system across a wide variety of organizational settings. While reinforcement learning will let the system learn from its environment through feedback and refine its detection strategies over time, federated learning will enable collaborative learning with the guarantee of data privacy.

Other improvement might be achieved by incorporating XAI methods into the system, as this could help enhance transparency and trust through clear explanations of decisions taken by threat detection models. This will be very beneficial for regulatory compliance and auditing.

Finally, the performance assessment of the system in adversarial settings and development of robust defenses against adversarial attacks would provide further evidence for real-world effectiveness. This direction allows for great potential to contribute toward next-generation Security Operation Centers and their capabilities in mitigating emerging cyber threats.

7 Conclusion

This research work has proposed a novel approach to NIDS by integrating the methods of Federated Learning, Graph Neural Networks, and Explainable AI techniques. The system here is designed to mitigate some critical challenges in traditional NIDS over scalability, privacy preservation, and interpretability. The results of this evaluation showed that the Federated GNN models-again, with emphasis on GAT architecture-relatively outperformed traditional deep learning models in multiclass network intrusion detection, including DDoS, Port Scan, Botnet, Brute Force, and Infiltration. With a Federated Learning framework, models could be collaboratively trained across decentralized environments without compromising data privacy while showing performance comparable to their centralized models. This approach allows the sensitive network data to stay within its local environment while still contributing to a robust global model.

The promising performance of the GNN models in capturing complex relational patterns of network traffic data was indicative of their future promise in the detection of sophisticated and evolving cyber threats utilizing these relationships. Among them, the GAT model showed especially good adaptability in complex network scenarios, as it is able to assign different attention weights to neighboring nodes. This greatly improves the detection rate. What's more, some techniques of XAI, such as SHAP and LIME, are integrated into the system and provide interpretable insights on how the model decides about its output. In such a way, the interpretability for the whole system is enhanced, which will lead to much more confidence from network administrators. This transparency is of paramount importance in practical applications, where the interpretation of a detected decision can inform effective threat mitigation and response.

These results have several implications regarding the future of network security. First, the proposed NIDS demonstrates that privacy-preserving machine learning methods, such as Federated Learning, can work with advanced models like GNNs to enhance detection capability while ensuring data security. It makes the solution scalable and secure to be deployed in anything, from enterprise-level networks up to critical infrastructure systems. Future studies can be done on integrating more advanced GNN architectures and optimization of federated learning algorithms, which give robustness and efficiency to the models. Second, further development of the XAI capabilities into more detailed and user-friendly visualizations could help in making the system more usable by the network operators and cybersecurity professionals. Another potential benefit with this approach is that it enables taking advantage of adaptation opportunities for other domains where secure, scalable, and interpretable machine learning solutions are in demand.

References

[1] A. Khraisat, I. Gondal, P. Vamplew, and J. Kamruzzaman, "Survey of intrusion detection systems: techniques, datasets and challenges," *Cybersecurity*, vol. 2, no. 1, p. 20, 2019, doi: 10.1186/s42400-019-0038-7.

[2] A. Khraisat and A. Alazab, "A critical review of intrusion detection systems in the internet of things: techniques, deployment strategy, validation strategy, attacks, public datasets and challenges," *Cybersecurity*, vol. 4, no. 1, p. 18, 2021, doi: 10.1186/s42400-021-00077-7.

[3] M. Sarhan, S. Layeghy, N. Moustafa, and M. Portmann, "Cyber Threat Intelligence Sharing Scheme Based on Federated Learning for Network Intrusion Detection," *J. Netw. Syst. Manag.*, vol. 31, no. 1, p. 3, 2022, doi: 10.1007/s10922-022-09691-3.

[4] X. Gu, F. Sabrina, Z. Fan, and S. Sohail, "A Review of Privacy Enhancement Methods for Federated Learning in Healthcare Systems.," *Int. J. Environ. Res. Public Health*, vol. 20, no. 15, Aug. 2023, doi: 10.3390/ijerph20156539.

[5] S. Agrawal *et al.*, "Federated Learning for intrusion detection system: Concepts, challenges and future directions," *Comput. Commun.*, vol. 195, pp. 346–361, 2022, doi: 10.1016/j.comcom.2022.09.012.

[6] M. Bhavsar, K. Roy, J. Kelly, and O. Olusola, "Anomaly-based intrusion detection system for IoT application," *Discov. Internet Things*, vol. 3, no. 1, p. 5, 2023, doi: 10.1007/s43926-023-00034-5.

[7] R. Mohammad, F. Saeed, A. A. Almazroi, F. S. Alsubaei, and A. A.

Almazroi, "Enhancing Intrusion Detection Systems Using a Deep Learning and Data Augmentation Approach," *Systems*, vol. 12, no. 3. 2024, doi: 10.3390/systems12030079.

[8] N. Rieke *et al.*, "The future of digital health with federated learning.," *NPJ Digit. Med.*, vol. 3, p. 119, 2020, doi: 10.1038/s41746-020-00323-1.

[9] S. Rahmani, A. Baghbani, N. Bouguila, and Z. Patterson, "Graph Neural Networks for Intelligent Transportation Systems: A Survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 8, pp. 8846–8885, 2023, doi: 10.1109/TITS.2023.3257759.

[10] F. Charmet *et al.*, "Explainable artificial intelligence for cybersecurity: a literature survey," *Ann. Telecommun.*, vol. 77, no. 11, pp. 789–812, 2022, doi: 10.1007/s12243-022-00926-7.

[11] A. Nadeem *et al.*, "SoK: Explainable Machine Learning for Computer Security Applications," *Proc. - 8th IEEE Eur. Symp. Secur. Privacy, Euro S P 2023*, pp. 221–240, 2023, doi: 10.1109/EuroSP57164.2023.00022.

[12] M. Ahmad, Q. Riaz, M. Zeeshan, H. Tahir, S. A. Haider, and M. S. Khan, "Intrusion detection in internet of things using supervised machine learning based on application and transport layer features using UNSW-NB15 data-set," *EURASIP J. Wirel. Commun. Netw.*, vol. 2021, no. 1, p. 10, 2021, doi: 10.1186/s13638-021-01893-8.

[13] J. Zhou *et al.*, "Graph neural networks: A review of methods and applications," *AI Open*, vol. 1, pp. 57–81, 2020, doi: https://doi.org/10.1016/j.aiopen.2021.01.001.

[14] J. L. Hernandez-Ramos *et al.*, "Intrusion Detection based on Federated Learning: a systematic review," 2023, [Online]. Available: http://arxiv.org/abs/2308.09522.

[15] E. Ntizikira, W. Lei, F. Alblehai, K. Saleem, and M. A. Lodhi, "Secure and Privacy-Preserving Intrusion Detection and Prevention in the Internet of Unmanned Aerial Vehicles," *Sensors*, vol. 23, no. 19. 2023, doi: 10.3390/s23198077.

[16] C. I. Nwakanma *et al.*, "Explainable Artificial Intelligence (XAI) for Intrusion Detection and Mitigation in Intelligent Connected Vehicles: A Review," *Applied Sciences*, vol. 13, no. 3. 2023, doi: 10.3390/app13031252.

[17] V. Vimbi, N. Shaffi, and M. Mahmud, "Interpreting artificial intelligence models: a systematic review on the application of LIME and SHAP in Alzheimer's disease detection.," *Brain informatics*, vol. 11, no. 1, p. 10, Apr. 2024, doi: 10.1186/s40708-024-00222-1.

[18] S. A. Allegri, K. McCoy, and C. S. Mitchell, "CompositeView: A Network-Based Visualization Tool.," *Big data Cogn. Comput.*, vol. 6, no. 2, Jun. 2022, doi: 10.3390/bdcc6020066.

[19] O. R. By and R. College, "NATIONAL CONFERENCE on VLSI , Publication Partner : IJARIIE RECOGNITION OF ADHD SYNDROME," vol. 4396, no. 1, 2023.

[20] L. Li, Y. Lu, G. Yang, and X. Yan, "End-to-End Network Intrusion Detection Based on Contrastive Learning.," *Sensors (Basel).*, vol. 24, no. 7, Mar. 2024, doi: 10.3390/s24072122.

[21] H. Noor, N. Islam, M. S. Hossain, N. Kamarudin, M. Raiaan, and S. Azam, *Determining the Optimal Number of GAT and GCN Layers for Node Classification in Graph Neural Networks*. 2023.

[22] E. U. Qazi, M. H. Faheem, and T. Zia, "HDLNIDS: Hybrid Deep-Learning-Based Network Intrusion Detection System," *Applied Sciences*, vol. 13, no. 8. 2023, doi: 10.3390/app13084921.

[23] M. Keshk, N. Koroniotis, N. Pham, N. Moustafa, B. Turnbull, and A. Y. Zomaya, "An explainable deep learning-enabled intrusion detection framework in IoT networks," *Inf. Sci. (Ny).*, vol. 639, p. 119000, 2023, doi: https://doi.org/10.1016/j.ins.2023.119000.

[24] R. Hamilton and P. Papadopoulos, *Using SHAP Values and Machine Learning to Understand Trends in the Transient Stability Limit*. 2023.