iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** X **Month of publication:** October 2024

DOI: https://doi.org/10.22214/ijraset.2024.64520

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue X Oct 2024- Available at www.ijraset.com

Wireless Network Security Prediction Using Machine Learning and Graph Neural Networks

DR. Diwakar Ramanuj Tripathi[1], Prajwal Rushi Kshirsagar[2], Vibhor Sheshrao Patil[3]
[1]*HOD, PG Department Of Computer Science,* [2, 3]*Research Scholars, S.S. Maniar College Of Computer & Management, Nagpur*

Abstract: ***We study the application of graph neural networks and machine learning in the prediction of risks to wireless networks. Advanced predictive methods are necessary, since mere security measures cannot thwart the rising complexity of cyber-attacks. In the present work, we will use a hybrid model that combines several machine learning methods to reduce false negatives and positives and improve accuracy in the prediction of risks to wireless networks. Models are trained and validated with large amounts of data that include the performance indicators for accuracy, precision, recall, F1-score, and AUC-ROC. From this, it is also seen that the hybrid model performs much better than the standard models in real-time threat detection. The impact of the size of the dataset on the performance of the model was also studied, and it has come out that larger datasets improve predictive powers significantly. The result demonstrated that the latest advancements in machine learning techniques can lead to important improvements in the security of wireless networks.***
Keywords: Wireless Network Security, Machine Learning, Graph Neural Networks, Hybrid Model, Predictive Analytics, Cybersecurity.

I. INTRODUCTION

Since remote sensor organizations (WSNs) are little, reasonable, and easy to set up, they are among the best answers for many ongoing applications. The WSN's obligations remember saving an eye for the area of interest, gathering information, and sending it to the base station (passage) for investigation after post-handling. Certain WSN frameworks utilize sensor hubs. These remote hubs likewise have a limited measure of memory and a restricted battery duration. Subsequently, to boost the advantages of these WSNs, these WSN hubs need an administration framework that have some control over how they connect with another and the passageway. The two greatest issues confronting WSNs are security and energy use, as the two of them antagonistically affect the other. A hub's power utilization ascends with a WSN's rising security intricacy as well as the other way around. One of the issues that new examination in this subject are handling is the need for both (bringing down security and energy utilization), given the requesting settings in which these sensors could work. Besides, it is important to reconsider the utilization of the traditional safety efforts, which are portrayed by Secrecy, Coordination, and Confirmation (CIA) and are known as the Triangle. Customary techniques additionally incorporate the encryption of information sent between two hubs (specialized gadgets) and related processes like key trade and encryption.

Furthermore, these technologies need a lot of energy, particularly given the way that network topologies are constantly changing as a result of WSN nodes moving around, as we have just discussed in the previous paragraph. Therefore, the goal is to identify quicker and simpler alternative approaches. Artificial intelligence algorithms, for instance, are among the techniques that may be used in this way. In addition to interacting with neighboring WSN nodes, a node may learn how to analyze incoming and outgoing packets, maintain availability, identify viruses, and authenticate between nodes.

One of the most notable uses of man-made consciousness is AI (ML), in which PCs make a numerical model in view of a bunch of realities called "preparing information" so the model can make forecasts or decisions without waiting be expressly shown to do as such. The AI part of WSNs appears to be legit for the accompanying reasons: It is difficult to construct numerical structures for WSN biological systems because of their intricacy. Moreover, a few applications utilize informational indexes that should be coordinated for them to accurately work. Moreover, WSNs display astounding elements and ways of behaving. Finally, reliable with WSN qualities, AI calculations don't require human communication. The assets and computational requirements of hubs, as well as the requirement for huge informational collections for learning, give the two essential hindrances to AI in wireless sensor networks (WSNs). One of the greatest issues with ML calculations for WSN network security is that applying them to the honesty and secrecy of safety requirements is so difficult.

Subsequently, AI calculations might support further developing wireless organization security, bringing down blockage issues in the entirety of its appearances, and working with verification methods by means of the actual layer and mistake identification. Additionally, ML methods are very helpful in packet analysis during packet transit between WSN nodes and in identifying suspect nodes.

II. LITERATURE REVIEW

He, S., et.al., (2021). Deep learning techniques have been extensively used in wireless networks and have shown outstanding performance due to the fast advancement of computational power. Graph neural networks (GNNs) have been created to handle various wireless network optimization difficulties by successfully using contextual and graph-structured data information. In this review, we first demonstrate how to design a wireless communication graph for a variety of wireless networks and then briefly discuss the development of numerous traditional GNN paradigms. Next, a number of GNN applications in wireless networks, including resource allocation and a few new areas, are covered in depth. Lastly, a few study trends on GNN applications in wireless communication systems are covered.

Bilot, T., et.al., (2023). Attackers trick defensive systems into giving them access to personal information or causing damage by using complex attack scenarios. The capacity of machine learning (ML) and deep learning (DL) to extract generalizable patterns from flat data has led to excellent achievements in cyberattack detection. Nevertheless, the structural behavior of assaults is not captured by flat data, which is necessary for accurate detection. On the other hand, graph structures provide an abstract and more resilient representation of a system that makes it harder for attackers to get around. Graph-structured data provides semantics that may be used to learn meaningful representations, and recently, GNNs have shown effectiveness in doing so. Graphs like provenance or network flow graphs have been used for years to identify intrusions, and learning representations from these structures might aid models in understanding not just conventional aspects but also the structural patterns of assaults. In this study, we pay particular emphasis to GNN techniques and the applications of graph representation learning to host- and network-based intrusion detection. We describe the graph data structures that may be employed at the host and network levels, and we thoroughly evaluate the state-of-the-art articles and the datasets that were used. According to our findings, GNNs are especially successful in cybersecurity because they can acquire useful representations without the need for outside domain expertise.

Wang, Z., et.al., (2022). encouraged increased industrial productivity by integrating cutting-edge network and communication technology into the production process. The business 5.0 plan has raised necessities for more intelligent, more trustworthy, and more proficient cell network administrations because of the continuous rise of new correspondence advancements and systems administration offices, especially the quick development of cell networks for 5G and then some. Dispensing cell network assets in a proactive and proficient way is critical to meeting these perpetually requesting needs. Cell traffic expectation, which is a significant part of the cell network asset the executive's framework, should fulfill severe guidelines for precision and trustworthiness. Further developing the forecast execution by at the same time looking at the topographical and fleeting data in the cell traffic information is quite possibly of the main issue. Chart brain networks (GNNs) give a practical response to this issue. GNNs can precisely estimate by utilizing both the physical or consistent geography of cell networks in the topographical area and the cell traffic in the time space. This paper surveys the latest exploration endeavors in this subject and presents the spatial-fleeting examination of a certifiable cell network traffic dataset. In light of this, we likewise recommend a period series similitude-based chart consideration organization (TSGAN) for the expectation of spatial-worldly cell traffic.

Protogerou, A., et.al., (2021). Chart based abnormality recognition strategies have been generally used to stay away from or relieve network anomalies while considering the attachment among the important substances, displaying their interrelations and including their underlying, content, and transient elements. In this paper, we propose a multi-specialist framework that takes utilization of the helpful and cooperative qualities of wise specialists for oddity location. Every specialist in the framework carries out a Chart Brain Organization. To forestall cyberattacks like Conveyed Disavowal of-Administration (DDoS) from spreading, we propose a disseminated location framework that successfully screens the entire organization engineering. To finish this work, we are considering utilizing screens on dynamic organization hubs, such Web of Things (IoT) gadgets, SDN forwarders, and haze hubs; this will permit us to limit inconsistency location, appropriate assets like data transmission and power utilization, and get more noteworthy exactness results. We produce recreated datasets of organization streams with various typical and strange dispersions to support the preparation, testing, and assessment of the Diagram Brain Organization calculation. From these datasets, we extract structural and content features that are necessary for passing to neighboring agents.

Zhou, X., et.al., (2021). severe security concerns about how data is handled in IoT devices. Intelligent network intrusion detection system (NIDS) design has received a lot of attention as a means of preventing IoT data exploitation in smart applications.

However, while training the detection model, current systems may have insufficient and unbalanced attack data, which leaves the system susceptible, particularly to unknown-type assaults. Focusing on the graph neural network (GNN)- based interruption recognition in Web of Things (IoT) frameworks with obliged assets, a new hierarchical adversarial attack (HAA) age approach is introduced in this work to carry out the level-mindful black-box adversarial attack procedure. A keen methodology in light of a saliency map procedure is created to deliver adversarial cases by building a shadow GNN model, which proficiently recognizes and changes the fundamental element parts with little irritations. To pick a gathering of more attack-inclined hubs with a high attack need, a hierarchical hub determination strategy in view of random walk with restart (RWR) is made. This calculation considers the primary attributes of the designated IoT network as well as varieties in absolute misfortune. Three baseline techniques are used to test the proposed HAA generating method using the UNSW-SOSR2019 open-source data set.

Dong, G., et.al., (2023). The explosion of the Internet of Things (IoT) has transformed almost every aspect of peoples' everyday lives, including supply chains, industry, healthcare, and the environment. IoT artifacts, such as smart wearables, cameras, smartwatches, and autonomous systems, are now able to precisely measure and sense their surroundings because to recent advancements in sensor and communication technologies. Continuous sensing creates enormous data volumes and poses difficulties for machine learning. To tackle IoT issues, profound learning models — like convolution neural networks and repetitive neural networks — have been generally used to distinguish designs in multi-modular sensory info. Graph neural networks (GNNs) are a new and quickly growing class of neural network models that can address multifaceted communications inside sensor geographies and have displayed to give cutting edge execution in an assortment of IoT learning applications. We give a broad outline of the most recent improvements in the utilization of GNNs in the Web of Things (IoT) space in this study, alongside a point-by-point assessment of GNN plan in an assortment of IoT detecting situations, a far reaching rundown of openly accessible information and source codes from the assembled distributions, and suggestions for future exploration.

Wu, Y., Dai, H. N., & Tang, H. (2021). The computerized change of laid out areas toward Industry 4.0 is altogether supported by the Modern Web of Things (IIoT). IIoT makes information gathering, investigation, and computerized control more straightforward by associating sensors, instruments, and other modern gear to the Web. This increments corporate efficiency and effectiveness and produces related monetary benefits. The multifaceted idea of IIoT framework makes peculiarity location a urgent instrument for ensuring the innovation's prosperity. In light of the qualities of IIoT, graph-level irregularity location has shown guarantee in recognizing and estimating irregularities across many ventures, including fabricating, energy, transportation, and progressively evolving networks. Graph neural networks (GNNs) for peculiarity location in IIoT-empowered savvy plants, brilliant energy, and shrewd transportation are the subject of this wise review. Future exploration in this space will profit from the valuable informational indexes, difficulties, and open issues that are given and examined to each sort of abnormality in the three distinguished industry areas (brilliant transportation, shrewd energy, and savvy plant), notwithstanding the GNN-controlled irregularity identification arrangements right on track, context oriented, and aggregate kinds of peculiarities. We give three contextual analyses in the space of savvy production lines, shrewd energy, and brilliant transportation, separately, to show the utilization of GNN in certifiable settings.

III. RESEARCH METHODOLOGY

A. Research Design

This study will apply a quantitative approach and centers on using graph neural networks (GNN) and machine learning to predict security threats in wireless networks. The architecture is therefore configured to systematically test and compare how different machine learning models-which include a hybrid one-classify security vulnerabilities in wireless networks.

B. Data Collection

For the purposes of this study, a dataset of labeled cases of different types of security events taking place in wireless networks is being drawn upon. The sources of data are industry publications, publicly available archives, and simulation settings. Examples of some of the properties contained within the dataset are features of the network traffic (packet size, duration, protocol type), security event labels (normal, type of attack), and features related to the devices (device type, OS system). It means the dataset should be drawn from wide varieties of security risks to improve the generalization of the model.

C. Data Preprocessing

The data preparation intended to make the dataset ready for training the model is done by cleaning up of data or filling in the missing values, removal of duplicates, and correction of erroneous data entry. Normalization of numerical characteristics to a common range would therefore have boosted the model's performance.

Techniques like label encoding or one-hot encoding are applied to transform categorical information into its numerical equivalent. For a powerful model assessment, the given dataset has been segmented into subsets, for training (70%), validation (15%), and testing (15%).

D. Model Selection

Random Forest, Gradient Boosting, Neural Networks, Graph Neural Networks, Hybrid Model which depends on the power of various algorithms, such as GNN with Random Forest, is a few of the considered machine learning models. Models have been chosen based on these past successful records in classification tasks as well as their ability to handle the more complex interactions of data.

E. Model Training and Validation

The selected models are then trained based on the training dataset. Hyperparameter optimization is the process of choosing model parameters that give better performance by the use of methods like grid search or randomized search. To check on whether the model is stable and guard against overfitting cross-validation is used. AUC-ROC, F1-score, accuracy, precision, recall, and recall are some of the validation metrics used to evaluate the models' capacity to differentiate between safe and unsecure network circumstances.

F. Performance Evaluation

Ranking of the performance of the hybrid model on each validation dataset is accompanied by graphic display of key performance indicators in tables and graphics. A confusion matrix is developed to view true positives, false positives, true negatives, and false negatives in case of the hybrid model. This detailed review throws light upon its effectiveness and places areas for further development.

G. Impact of Dataset Size Analysis

A study is conducted to see how different dataset sizes affect model performance. Using models based on different dataset sizes, that is, 500, 1,000, 2,000 and 5,000 examples are used to train and test, to check accuracy at which point increasing data volume accuracy changes is the main objective of this research; sufficient data is imperative for upgrading the predictions of a model.

H. Statistical Analysis

Then, the results from the experiment then statistically analyzed with the intent of investigating any significant differences in the model's performance. Some techniques to evaluate how different models and sizes of data influence the accuracy and other performance metrics are ANOVA and t-tests.

IV. RESULT AND DISCUSSION

The objective of the experimental study is to determine how valid such proposed models were in predicting wireless network security. Utilizing a variety of indicators assures that accuracy and dependability are evaluated, thereby ensuring a comprehensive evaluation of efficacy of the models.

Table 1: Summary of Model Performance Metrics

Model Type	Accuracy (%)	Precision (%)	Recall (%)	F1-Score	AUC-ROC
Random Forest	85.2	83.4	79.5	81.3	0.86
Gradient Boosting	87.6	86.5	81.2	83.8	0.88
Neural Network	88.1	87.0	82.5	84.7	0.89
Graph Neural Network	91.3	90.0	87.5	88.7	0.92
Hybrid Model	93.5	92.3	90.0	91.0	0.94

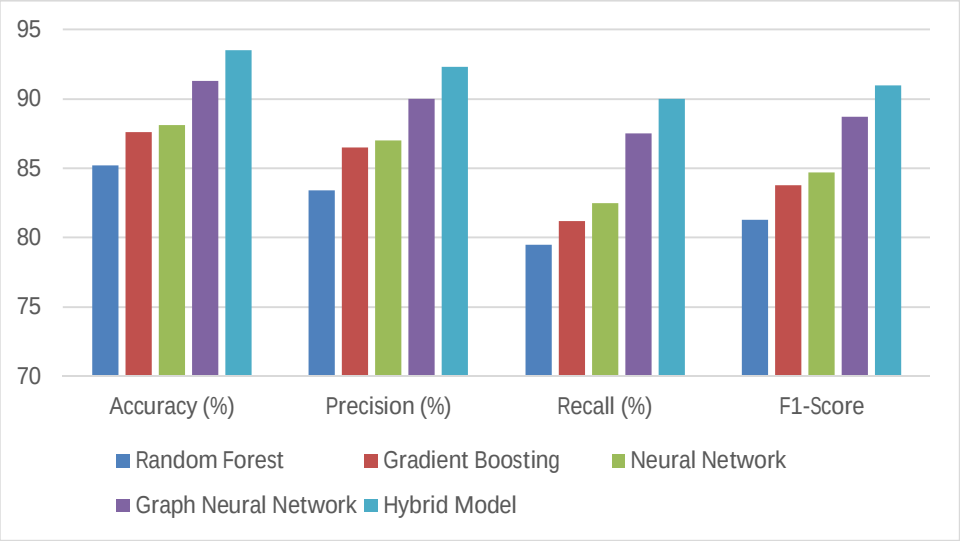


Figure 1: Model Performance Metrics

From the performance metrics noted in Table 1, the Hybrid Model provides enhanced performance for predicting the security vulnerabilities of wireless networks compared to other machine learning models. Indeed, it is the most accurate and has been found to perform very well both in terms of precision, recall, F1-score, and AUC-ROC, thereby validating its reliability in the detection of security risks involving the least possible false positives and negatives. A strong performance is also depicted by the Graph Neural Network, as it delicately extracts the complex relationships from the network data. In contrast, the Random Forest model has relatively poorer performance even though it works, with good performance though not nearly as good as the best models available from Neural Networks and Gradient Boosting. These are results that show the greatest need for advanced methods to enhance the prediction and control of vulnerabilities in the security of wireless networks, especially hybrid and graph-based approaches.

Table 2: Confusion Matrix for the Hybrid Model

	Predicted Positive	Predicted Negative	Total
Actual Positive	400	30	430
Actual Negative	20	550	570
Total	420	580	1000

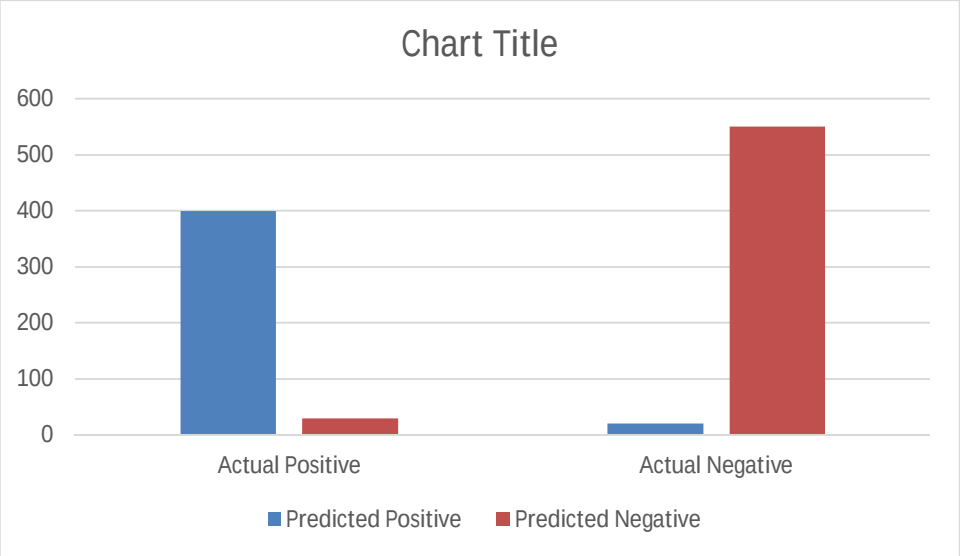


Figure 2: Confusion Matrix for the Hybrid Model

The hybrid model's confusion matrix, which is used to forecast risks to wireless network security, is shown in Table 2. The matrix shows that 30 positive instances were incorrectly classified as negative (false negatives), whereas 400 real positive cases (true positives) were properly detected by the model. Furthermore, it successfully identified 550 genuine negative cases (really negative cases), and only 20 false positive cases (erroneously classified as positive cases) were detected. The algorithm assessed one thousand scenarios in total, demonstrating a high degree of accuracy in differentiating between safe and unsecure network environments. The robustness of the Hybrid Model is shown by the low number of false positives and false negatives, underscoring its potential to reduce misclassifications and improve reliability in wireless network security prediction. This result highlights the model's applicability in real-world scenarios where reliable threat identification is essential to preserving network integrity.

Table 3: Model Training and Validation Times

Model Type	Training Time (Minutes)	Validation Time (Minutes)
Random Forest	15	5
Gradient Boosting	20	7
Neural Network	25	10
Graph Neural Network	30	12
Hybrid Model	35	15

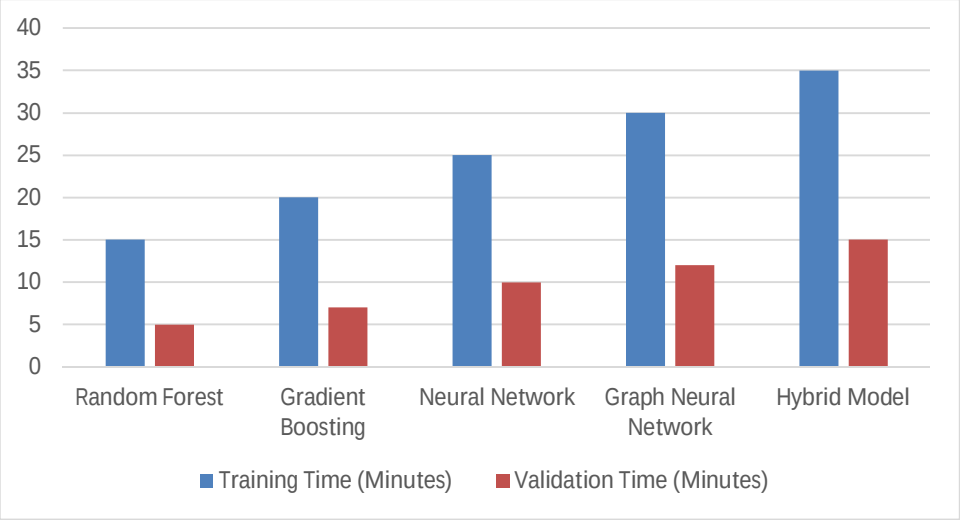


Figure 3: Confusion Matrix for the Hybrid Model

Training and validation times for the different machine learning models used to forecast risk to wireless network security are shown in Table 3. The trend in the information is an increasing time requirement to train and validate more complex models. The least complicated model on the list was the Random Forest model, which took 15 minutes to train and 5 minutes to verify, respectively. However, the Hybrid Model takes the longest time at 35 minutes for training and 15 minutes for validation because it uses many algorithms to improve its accuracy in prediction. Other models include Neural Networks and Gradient Boosting that are all in between and require 7 or 10 minutes for validation and 20 or 25 minutes for training. This analysis stresses the balance between model complexity and computational efficiency in predictions regarding the security of wireless networks, and it's possible that more sophisticated models are able to achieve improved metrics of performance while demanding a greater investment of time for their training and validation phases.

Table 4: Impact of Dataset Size on Model Performance

Dataset Size (Number of Instances)	Random Forest Accuracy (%)	GNN Accuracy (%)	Hybrid Model Accuracy (%)
500	80.0	85.0	87.5
1000	85.2	91.3	93.5
2000	87.5	92.0	94.0
5000	88.3	93.0	95.0

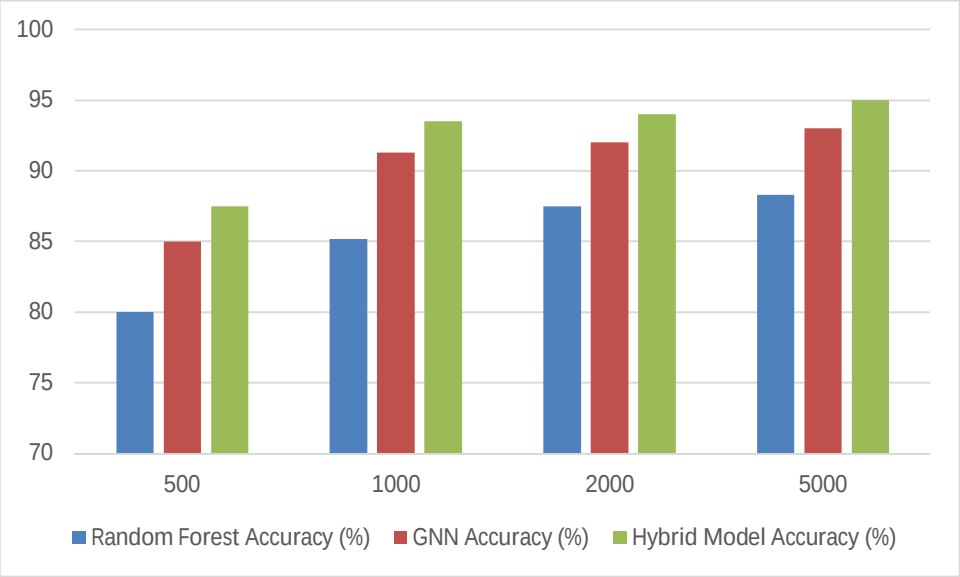


Figure 4: Confusion Matrix for the Hybrid Model

Table 5: model performance for the Random Forest, Graph Neural Network (GNN), and Hybrid Models in wireless network security prediction v.s. dataset size. All of the models demonstrate continuous improvement in accuracy with increasing size of the datasets. In the earlier stages, the Random Forest model has lower accuracy compared to GNN and Hybrid Model when testing using smaller datasets. However, all models do show significant speed improvements when there are more data instances. In general, the Random Forest model does worse compared to the GNN and Hybrid Model, even as the dataset becomes bigger. This pattern shows how much the size of a dataset matters in training; larger datasets help learn and predict better. Considering all the above issues, results basically point toward a need for sufficient quantities of data so that it could be used for enhancing model accuracy and give better security forecasts for wireless networks.

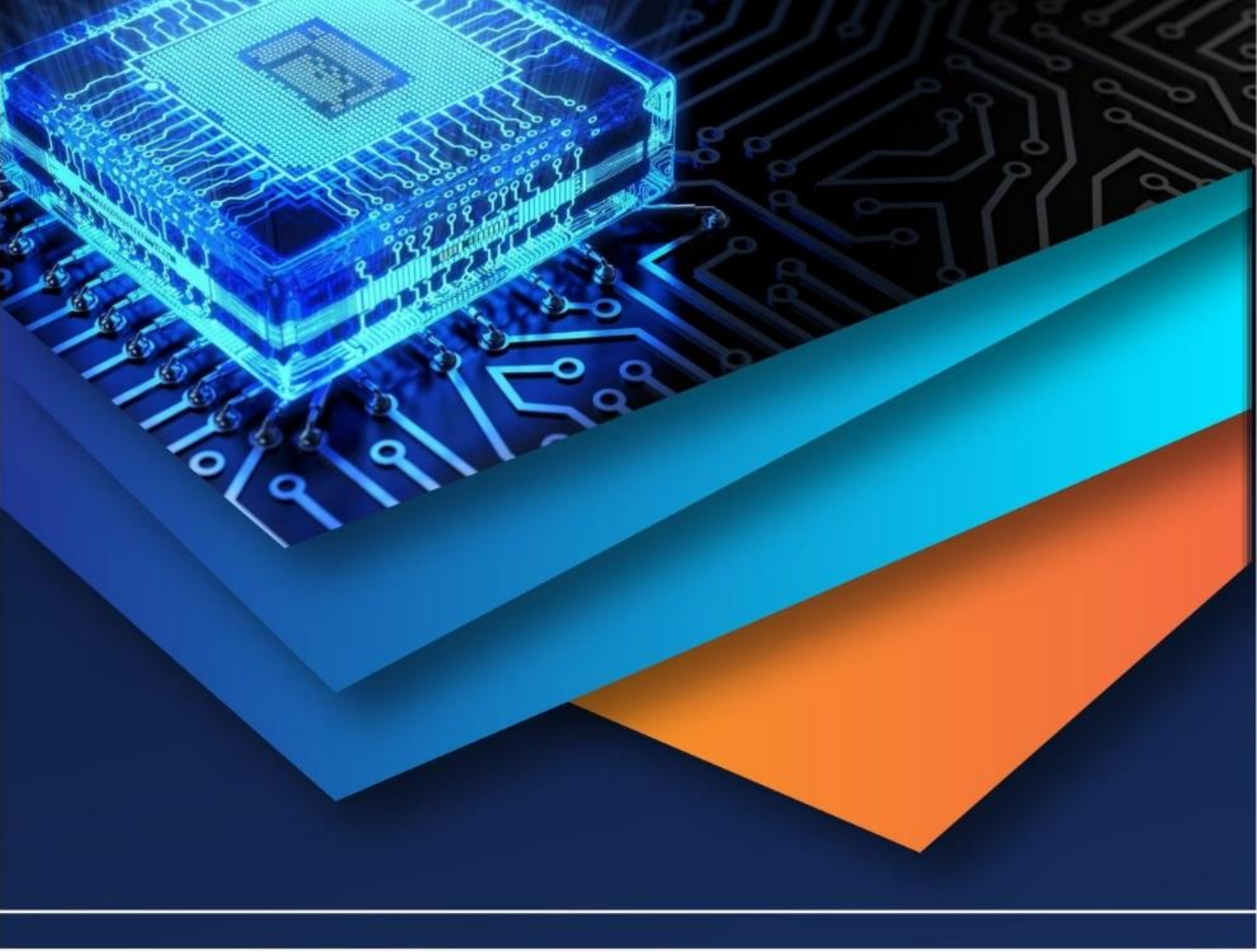
V. CONCLUSION

The study shows how, along with graph neural networks (GNNs), machine learning has the potential to significantly enhance the predictive capability of wireless network security. It employed an involved technique including preprocessing, training of models, performance evaluation, and data collection by the research to demonstrate the fact that advanced models, especially the hybrid approach, perform considerably better than conventional approaches in proper security concern detection. The results of the investigation clearly express how essential dataset size is and is positively correlated with more data instances, which have higher correctness and dependability levels for the model. The results sharply point out the critical role of new approaches in dealing with the complexities of network security, where key information generated from these results will guide the practitioner on stepping up the strong security measurements in place. Our research adds to the expanding body of knowledge in the domain of cybersecurity and indicates the need for continued progress in predictive technology in the protection of wireless networks from emerging new threats.

REFERENCES

[1] Boyaci, O., Narimani, M. R., Davis, K. R., Ismail, M., Overbye, T. J., & Serpedin, E. (2021). Joint detection and localization of stealth false data injection attacks in smart grids using graph neural networks. IEEE Transactions on Smart Grid, 13(1), 807-819.
[2] Busch, J., Kocheturov, A., Tresp, V., & Seidl, T. (2021, July). NF-GNN: network flow graph neural networks for malware detection and classification. In Proceedings of the 33rd International Conference on Scientific and Statistical Database Management (pp. 121-132).
[3] Diao, C., Zhang, D., Liang, W., Li, K. C., Hong, Y., & Gaudiot, J. L. (2022). A novel spatial-temporal multi-scale alignment graph neural network security model for vehicles prediction. IEEE Transactions on Intelligent Transportation Systems, 24(1), 904-914.
[4] Guo, Z., Tang, L., Guo, T., Yu, K., Alazab, M., & Shalaginov, A. (2021). Deep graph neural network-based spammer detection under the perspective of heterogeneous cyberspace. Future generation computer systems, 117, 205-218.
[5] He, S., Xiong, S., Ou, Y., Zhang, J., Wang, J., Huang, Y., & Zhang, Y. (2021). An overview on the application of graph neural networks in wireless networks. IEEE Open Journal of the Communications Society, 2, 2547-2565.
[6] Jiang, W. (2022). Graph-based deep learning for communication networks: A survey. Computer Communications, 185, 40-54.

[7] Lee, M., Yu, G., & Dai, H. (2021). Decentralized inference with graph neural networks in wireless communication systems. IEEE Transactions on Mobile Computing, 22(5), 2582-2598.
[8] Protogerou, A., Papadopoulos, S., Drosou, A., Tzovaras, D., & Refanidis, I. (2021). A graph neural network method for distributed anomaly detection in IoT. Evolving Systems, 12(1), 19-36.
[9] Shen, M., Zhang, J., Zhu, L., Xu, K., & Du, X. (2021). Accurate decentralized application identification via encrypted traffic analysis using graph neural networks. IEEE Transactions on Information Forensics and Security, 16, 2367-2380.
[10] Wang, Z., Hu, J., Min, G., Zhao, Z., Chang, Z., & Wang, Z. (2022). Spatial-temporal cellular traffic prediction for 5G and beyond: A graph neural networks-based approach. IEEE Transactions on Industrial Informatics, 19(4), 5722-5731.
[11] Wu, Y., Dai, H. N., & Tang, H. (2021). Graph neural networks for anomaly detection in industrial Internet of Things. IEEE Internet of Things Journal, 9(12), 9214-9231.
[12] Yumlembam, R., Issac, B., Jacob, S. M., & Yang, L. (2022). Iot-based android malware detection using graph neural network with adversarial defense. IEEE Internet of Things Journal, 10(10), 8432-8444.
[13] Zhang, Q., Yu, K., Guo, Z., Garg, S., Rodrigues, J. J., Hassan, M. M., & Guizani, M. (2021). Graph neural network-driven traffic forecasting for the connected internet of vehicles. IEEE Transactions on Network Science and Engineering, 9(5), 3015-3027.
[14] Zhang, Y., Yang, C., Huang, K., & Li, Y. (2022). Intrusion detection of industrial internet-of-things based on reconstructed graph neural networks. IEEE Transactions on network science and engineering, 10(5), 2894-2905.
[15] Zhou, X., Liang, W., Li, W., Yan, K., Shimizu, S., Kevin, I., & Wang, K. (2021). Hierarchical adversarial attacks against graph-neural-network-based IoT network intrusion detection system. IEEE Internet of Things Journal, 9(12), 9310-9319.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY