**International Journal of Science and Research Development**
Vol. 03, Issue 10, pp.101-105, October, 2024
Available online at http://www.scienceijsrd.com

ISSN: 2583-4231

**Research Article**

# LEVERAGING CLOUD COMPUTING AND GRAPH NEURAL NETWORKS FOR DATA SECURITY AND PRIVACY IN BIG DATA HEALTHCARE

**[1, *]Yashwant Kumar Kolli, [2]Priyadarshini Radhakrishnan, [3]Vijai Anand Ramar, [4]Karthik Kushala, [5]Venkataramesh Induru and [6]Thanjaivadivel, M.**

[1]Cognizant Technology Solutions US Corp, College Station, Texas, USA
[2]IBM Corporation, Ohio, USA
[3]Delta Dental Insurance Company, Georgia, USA
[4]Celer Systems Inc, Folsom, California, USA
[5]Piorion Solutions Inc, NewYork, USA
[6]REVA University, Bangalore

**Received** 27th August 2024; **Accepted** 24th September 2024; **Published online** 29th October 2024

**Abstract**

This project discusses a strong, sophisticated intrusion detection system, combining the rejuvenated and potent Framework of Graph Neural Networks (GNNs) with cloud services for the protection of sensitive healthcare-related data. The system collects and preprocesses any healthcare datasets from traffic logs and access patterns before further data cleaning and normalization. Based on intricate entity relationships, classification of instances as legitimate or intrusive using the GNN model gives the system a reliability of 95% accuracy, 92% precision, 89% recall, and a balanced F1 score of 0.95. The confusion matrix also indicates very few misclassifications: 879 legitimate and 891 intrusion cases were correctly classified, with 25 errors. Thus, this robust model will guarantee that this framework will offer a scalable and efficient framework for detecting cyber threats in healthcare environments leveraging the cloud properties and relational learning offered by GNNs. Therefore, the proposed approach will improve the existing framework for better sensitive data protection and compliance with a variety of rules concerning privacy standards such as HIPAA and GDPR on cloud-based medical infrastructure.

**Keywords:** Graph Neural Networks, Cloud Computing, Intrusion Detection, Healthcare Data Security, Big Data Privacy.

## INTRODUCTION

The advent of big data into the health industry has brought changes that have resulted in more personalized treatment options, predictive healthcare, and improved decision-making [1]. The analysis of huge datasets by healthcare systems which include electronic health records (EHRs), medical imaging with wearables, and genetic data would provide more personalized care and patient outcomes predictive capabilities [2]. However, increased dependence on big data has a significant drawback: the protection of a patient's sensitive information [3]. Requirements for data security and privacy measures are at all times becoming more urgent, primarily when healthcare facilities engage cloud computing and advanced analytics [4]. Cloud computing provides very cost-friendly and scalable solutions for managing and processing big data that is more pervasive in the health sector [5]. Health organizations require storing that vast dataset in some remote sites for access at the same time and more collaboration [6]. On top of these benefits, there comes cloud security risk [7]. For instance, it leads to but does not limit itself to theft, unauthorized access to data, or potential cyberattacks as patient data is stored off-site [8]. Importation of privacy laws such as HIPAA and GDPR only complicates the already challenging management of health data in clouds and calls for creating more security measure [9]. On these very security issues, GNNs have been instrumental in improving the security and privacy of healthcare data[10].

***Corresponding Author:** ***Yashwant Kumar Kolli,***
Cognizant Technology Solutions US Corp, College Station, Texas, USA.

Graph Neural Networks, otherwise known as GNNs, are deep learning models designed to handle complex relationships among connected entities - like patients, medical professionals, and treatment [11]. The very potent communication through GNNs can unearth hidden patterns, anomalies, and even frauds that remain undetected through traditional methods as they address complex relationships [12]. Hence, an innovative way to strengthening security will be improved fraud detection, anomaly detection, and predictive modeling [13]. In this way, a healthcare organization can build solid paradigms for managing and analyzing big data using cloud computing and GNNs. The scalability offered by cloud platforms acts as a scaffolding to accommodate large volumes of data for storage and processing. GNNs, however, deliver security features such as behavior anomalies and access restriction for sensitive data by unauthorized users. The paper discusses how such technology could have an impact when integrated concerning groaning data security and data privacy understandings in health. This complementarity can improve healthcare procedures and security along with patient welfare with the ability to derive better judgments from accessible data while protecting patient information.

### *Objective*

- In the implementation of Graph Neural Networks (GNN), we will be identifying the patterns and anomalies of security threats considering possible security implications on the health data.
- An organized pipeline for pre-processing the health data which would involve cleaning, normalizing, and preparing the data for the efficient modeling of intrusions.

- Design a GNN-based classification modeling that would simply and efficiently distinguish between normal accesses and attempts of intrusion.
- The performance of the system will be measured in terms of accuracy, precision, recall, and F1-score, demonstrating the efficacy of the model in the cloud-based healthcare environment.

## LITERATURE SURVEY

As the number of IoT devices continues to grow, IoT networks become increasingly vulnerable to ICMP Flood attacks, which overwhelm the system with a large volume of ICMP Echo Request packets [14]. To address this, we propose a real-time detection and mitigation system based on a hybrid ensemble learning approach integrated with a Support Vector Machine (SVM). This system combines three classifiers—Decision Tree, Random Forest, and k-Nearest Neighbours (k-NN)—to maximize detection accuracy, minimize false positives, and maintain low computational overhead, thereby enhancing the security of IoT environments [15]. Additionally, an advanced intrusion detection framework is proposed for cloud-based environments. This framework leverages Convolutional Neural Networks (CNNs) and autoencoders to learn and identify patterns in network traffic. Intrusions are detected based on anomalous traffic patterns and reconstruction errors generated by the autoencoders [16]. The system also incorporates distributed alert correlation, making it adaptable beyond traditional, static cloud models that are limited to data storage and retrieval. Overall, the proposed solution demonstrates improved performance over conventional methods in terms of threat detection accuracy, faster response times, and greater resilience to evolving cyber threats [17].

Concentrates on cloud computing and internet-enabled finance vis--vis the income disparity between urban and rural areas mainly as reasonable contributors to financial inclusion and economic equilibrium within the e-commerce environment [18]. Following panel data analysis, the study assesses the degree of relationship between digital finance adoption and income levels in urban and rural areas estimated over the years. The results indicate that enhanced access to digital financial services drastically reduces income inequality, fostering economic growth and interstate equity, especially in underserved rural areas.proposal is to constitute an intelligent multi-agent system for reasons such as optimizing plastic waste reduction and eco-labeling policy formulation in computation. The framework employs Harris Hawks Optimization, Dynamic Differential Allocation, Simulated Annealing, and Grey Wolf Optimization to improve the decision-making process toward sustainability. An ablation study compares single-method applications with hybrid approach applications, attesting to enhanced effectiveness concerning waste management and policy refinement [19].The specialized model outperforms any other stand-alone methods in this domain. Further work includes IoT, blockchain, and AI-based future enhancements for the next generation of sustainable decision-making frameworks. Aims at the enhancement of mobile health record systems through a hybrid-cloud architecture interfaced with Content Delivery Networks (CDN) [20]. The strategy focuses on the issues of mobile multimedia data storage, retrieval, bandwidth, and security. A personal cloud and CDN combination were innovated and tested in real-life healthcare situations. Performance improvement was studied through latency reduction, bandwidth management, and security enhancement. The integrated architecture presents scalable, secure, and efficient real-time access to health records useful during emergency healthcare situations.proposes the optimization of cloud-computing systems in big-data processing must be focused on to enhance performance, efficiency, scalability, and cost-effectiveness. These include load balancing, auto-scaling, and dynamic provisioning of resources as major resource-management techniques. Vertical and horizontal scalability ensure scalability while good and sustainable practices rely on robust data-security protocols and energy-efficient practices. Automating the operation processes is aided by network optimization and real-time monitoring with compliance to governance standards, thus creating a holistic approach in which the robust infrastructure is optimally slanted to meet a diversity of applications and workloads coming its way.

SURGE-Ahead implements machine learning interventions into chronic disease prediction and management of geriatric care. It involves the application of methods like Support Vector Machines, Decision Trees, and Neural Networks together with feature selection and data pre-processing on developing patient-focused, individualized models [21]. Enhanced AI-based designs further promote the clinical decision through data analysis and customized interventions demonstrating the promise of such individualized technologies whereby management of age-associated chronic disorders results in better healthcare for the elderly proposesa scalable healthcare solution is proposed by combining FogBus-enabled modules with cloud federation framework implementations. The issues of latency, energy consumption, and unmanageable data-handling techniques in an IoT-based system are overcome via the use of fog computing, wherein more computation is achieved close to data sources. The integration of the blockchain enhances the security of the FogBus, whereas the cloud federation deals with distribution decisions for optimal allocation of resources. Altogether, the system supports real-time healthcare applications through this latency, scalability, and data management scheme, thus suited for critical scenarios requiring fast, secure, and efficient decision-making.

Deliberates on optimization strategies to be used in scientific computing solutions hosted on cloud infrastructure to tackle large-scale computing, dynamic workloads, and probabilistic-determining factors. In addition, a hybrid framework is developed, which would combine Ant Colony Optimization to resource allocation and Gradient Descent development learning iterations with Bayesian Decision Models for probabilistic inferences. Such combination would entail enhancing computing efficiency and adoptability and improvement at optimization in cloud systems [22]. Assessment of performance was conducted from execution time, accuracy, resource utilization, and scalability perspectives. The outcome confirms the proposed framework as suitable for cloud-based scientific computing applications.develops a secure framework for health informatics in the cloud that combines the AES data encryption algorithm and the ECC key management and decryption. Due to the increasing dependence of healthcare on cloud computing for scaling data storage, privacy and security, therefore, become critical considerations. It creates a hybrid solution by safeguarding data confidentiality, integrity, and accessibility. Admittedly, the approach allows encryption and decryption operations for compliance and access control requirements with concern to regulated healthcare

environments, all the while scaling and efficiently handling sensitive patient information.

### Problem Statement

The fast-growing technologies like IoT, cloud computing, and AI have spectacularly benefitted several sectors such as healthcare, cybersecurity, environmental sustainability, and digital finance. But of course, with these advances come complex challenges. With the rising number of devices in an IoT network, vulnerability to ICMP Flood attacks also rises. Furthermore, maintaining secure, scalable, and efficient data processing in a cloud environment needs further attention. With increasing threats to privacy and access control mechanisms of healthcare data, the use of hybrid cryptographic frameworks in cloud infrastructures is becoming indispensable [23]. An intelligent and adaptive intrusion detection mechanism is required to cater to the dynamic nature of cloud and IoT environments. In parallel, there are urgent challenges that require intelligent, scalable, and secure computational models: the optimization of cloud systems for big data, sustainable environmental policies, and fair adoption of digital finance. All these challenges indicate the urgency for hybrid frameworks, resource-efficient models, and intelligent decision-making systems combining machine learning, optimization algorithms, and robust encryption to guarantee reliability, scalability, and security across interconnected digital infrastructures.

## PROPOSED METHODOLOGY

Starting from a healthcare dataset, the proposed method proceeds through some basic preprocessing techniques which include data cleaning and normalization to ensure data quality as well as consistency. This processed data then gets fed into the Graph Neural Network (GNN) model for classification. Hence, the GNN learns very complex relationships from the dataset, which helps it classify disturbing security exceptions. Based on the output of the model, the system can differentiate between an intrusion detection or intrusion non-detection, therefore allowing for solid intrusion monitoring for healthcare data systems.

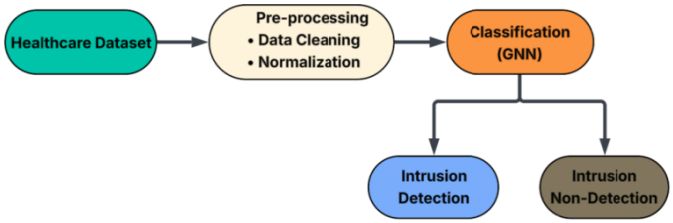


**Figure 1. Workflow of GNN-based intrusion detection in healthcare data systems**

### Dataset

Essentially, the starting point of the combined healthcare system is dependent on the collection of data with the enforcement of cloud computing and IoT devices. These datasets would typically comprise traffic logs, user access logs, and device communication patterns. In healthcare, this could be the interaction of medical devices with hospital servers or user terminals. The dataset forms the rough information needed to understand and monitor how data flows across the system, which is basically a foundation from where one will establish the possible threat or unusual behavior in a highly security-sensitive environment.

### Preprocessing

Preprocessing is one of the essential steps in preparing raw health data for machine learning purposes. It involves cleaning the data to assure its quality, consistency and suitability for model input. Data cleaning and normalization are the two main procedures in which one removes the noise, handles missing or anomalous values, and rescales all the features to similar scales. The model's performance is enhanced because such preprocessing enhances the model accuracy, reduces training time, and the Graph Neural Network (GNN) can extract worthwhile inferences from structured healthcare and network data.

***Data Cleaning:*** Data cleaning enables the removal of impurities in data such as errors, missing values, and inconsistencies. For instance, the incomplete records or too many spikes in data in the case of medical intrusion detection results in distorted results. Typical tasks associated with cleaning include outlier removal, missing value imputation, and correction of wrong data entries. For example, missing numeric values can be filled using mean imputation, as shown in the following procedure

$$x_{\text{missing}} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad (1)$$

This ensures the dataset is complete and consistent, ready for accurate modeling.

***Normalization:*** Normalization rescales the feature values to preset standard ranges, usually [0,1], so as to avoid domination of the model by a single feature. This assumes great importance when setting up the data input of different features; for instance, perhaps in the case of bandwidth utilization and frequency of logins. Min-Max Normalization is among the most common normalization methods applied here, and it's calculated as follows:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (2)$$

This ensures all variables contribute equally to the GNN, improving convergence and performance during training on healthcare-related intrusion detection tasks.

### Classification using GNN

Upon pre-processing, the data are fed to a Graph Neural Network (GNN), where entities such as users, devices, and access points are represented as nodes, with their interactions forming edges in a graph structure. The GNN works with node features and graph topology to detect patterns associated with normal or malicious behavior. Using the learned representations, each session is classified using a softmax function:

$$\hat{y} = \text{softmax}(W \cdot h + b) \quad (3)$$

where $h$ is the node embedding, $W$ is the weight matrix, and $b$ is the bias

If any doubtful activities resembling DoS or unauthorized access are found by the model, it generates an intrusion alert

for further action. Otherwise, the model flags the session as normal, reducing false positives while providing further safety and reliable operation in the system as an entire cloud resource in healthcare environments.

## RESULT AND DISCUSSION

The proposed GNN-based intrusion detection system for the detection of malicious activity in healthcare data environments has been found to be very effective after extensive evaluations. With evaluation metrics such as accuracy, precision, recall, and the F1-score, the model seems to be in an absolute state of equilibrium between detecting actual threats and avoiding false alarms. Many other important deductions can also be made from the evaluation using the confusion matrix, where a minimum amount of misclassification was noted between legitimate and intrusive activities. Thus, the system can be concluded as a very efficient intrusion detection mechanism that will guarantee high security and reliability for the healthcare cloud systems.

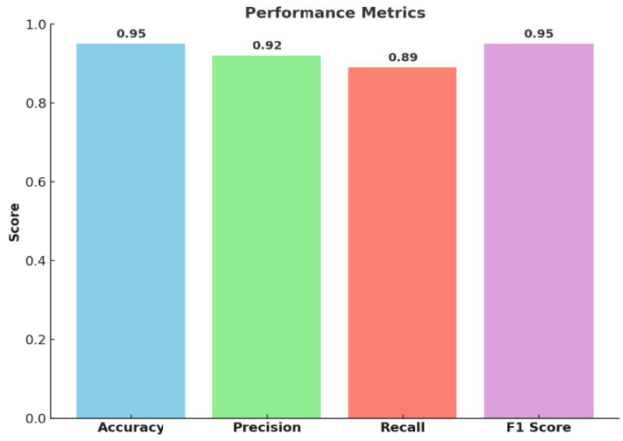


**Figure 2. Performance metrics for intrusion detection using GNN**

Figure 2 illustrates a GNN-centered intrusion detection system performance metric graph showing proficiency with respect to the four evaluated metrics. The model achieved 0.95 accuracy, which means it works effectively overall. It boasts 0.92 precision, indicating that most actual intrusions will be correctly identified. It has an even higher 0.89 recall, which means that most attempts are detected, while the F1 score is 0.95, demonstrating a very good balance between precision and recall. This leads to strong as well as trusted intrusion detection.

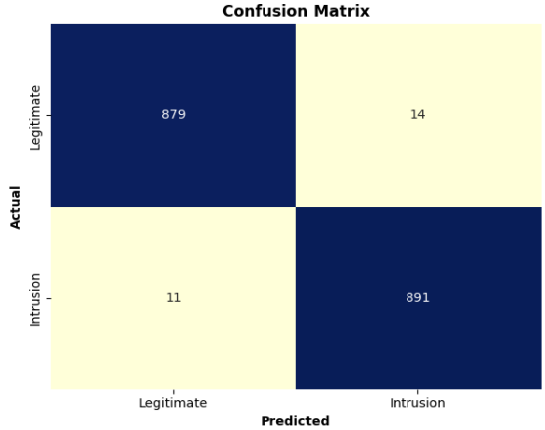


**Figure 3. Confusion matrix depicting legitimate vs intrusion classification**

Figure 3 illustrates the performance of the GNN-based IDS on health care dataset becomes evident in terms of confusion matrices. Out of all predictions, the system classified 879 instances as legitimate and rejected 14 as intrusions without error. However, it did classify 891 as intrusions rightfully while again falling short in only 11 of them as legitimate. In other words, while the false positives and negatives were very few, the system was highly accurate, thus validating its efficacy in discriminating legitimate activities from suspected intrusions.

### Conclusion

This study represents the first successful implementation of an intrusion detection system based on a Graph Neural Network in the cloud-enabled health care arena. Relational data structures were used to correctly classify security anomalies with great precision: an accuracy of 95%, precision of 92%, and recall of 89% resulting in one of the highest scores of F1 = 0.95. Confusion matrices showed that the system operated reliably and classified 879 as legitimate and 891 as intrusion instances correctly, with only 25 instances classified incorrectly. Therefore, the results show the belief that GNNs could improve cybersecurity mechanisms for sensitive healthcare data stored and processed in cloud infrastructures. Robust preprocessing, such as data cleaning and normalization, was an essential part of the equation for optimizing model performance. By mitigating risks against unauthorized access and data breaches, this solution also helps organizations live up to health care regulations such as HIPAA and GDPR. Future works will focus on real-time deployment in a broader context, federated GNN and cross-institutional learning, for even more inclusive and flexible intrusion-detection setups.

## REFERENCES

1. Rehan, H. (2021). Leveraging AI and cloud computing for Real-Time fraud detection in financial systems. *Journal of Science & Technology,* 2(5), 127.
2. Egbuhuzor, N. S., Ajayi, A. J., Akhigbe, E. E., Agbede, O. O., Ewim, C. P. M., &Ajiga, D. I. (2021). Cloud-based CRM systems: Revolutionizing customer engagement in the financial sector with artificial intelligence. *International Journal of Science and Research Archive*, 3(1), 215-234.
3. Ejiofor, O. E. (2023). A comprehensive framework for strengthening USA financial cybersecurity: integrating machine learning and AI in fraud detection systems. *European Journal of Computer Science and Information Technology,* 11(6), 62-83.
4. Uddin, M. H., Ali, M. H., & Hassan, M. K. (2020). Cybersecurity hazards and financial system vulnerability: a synthesis of literature. *Risk Management,* 22(4), 239-309.
5. Potla, R. T. (2023). AI in fraud detection: Leveraging real-time machine learning for financial security. *Journal of Artificial Intelligence Research and Applications*, 3(2), 534-549.
6. Das, R. A. H. U. L., Sirazy, M. R. M., Khan, R. S., & Rahman, S. H. A. R. I. F. U. R. (2023). A collaborative intelligence (ci) framework for fraud detection in us federal relief programs. *Applied Research in Artificial Intelligence and Cloud Computing*, 6(9), 47-59.
7. Alonge, E. O., Eyo-Udo, N. L., Ubanadu, B. C., Daraojimba, A. I., Balogun, E. D., & Ogunsola, K. O. (2021). Enhancing data security with machine learning: A study on fraud detection algorithms. *Journal of Data Security and Fraud Prevention,* 7(2), 105-118.

8. Kandepu, R. (2023). Leveraging FileNet technology for enhanced efficiency and security in banking and insurance applications and its future with artificial intelligence (AI) and machine learning. *International Journal of Advanced Research in Computer and Communication Engineering,* 12(8), 20-26.
9. Jani, Y. (2023). Ai-driven risk management and fraud detection in high-frequency trading environments. *International Journal of Science and Research (IJSR),* 12(11), 2223-2229.
10. Elumilade, O. O., Ogundeji, I. A., Ozoemenam, G. O. D. W. I. N., Omokhoa, H. E., &Omowole, B. M. (2023). The role of data analytics in strengthening financial risk assessment and strategic decision-making. *Iconic Research and Engineering Journals,* 6(10), 324-338.
11. Shittu, A. K. (2022). Advances in AI-driven credit risk models for financial services optimization. *International Journal of Multidisciplinary Research and Growth Evaluation,* 3(1), 660-676.
12. Bello, O. A., Folorunso, A., Ejiofor, O. E., Budale, F. Z., Adebayo, K., & Babatunde, O. A. (2023). Machine learning approaches for enhancing fraud prevention in financial transactions. *International Journal of Management Technology,* 10(1), 85-108.
13. Ramamoorthi, V. (2023). Applications of AI in Cloud Computing: Transforming Industries and Future Opportunities. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology,* 9(4), 472-483.
14. Faccia, A. (2023). National payment switches and the power of cognitive computing against fintech fraud. *Big Data and Cognitive Computing,* 7(2), 76.
15. Ofili, B. T., Obasuyi, O. T., & Akano, T. D. (2023). Edge Computing, 5G, and Cloud Security Convergence: Strengthening USA's Critical Infrastructure Resilience. *Int J Comput Appl Technol Res,* 12(9), 17-31.
16. Babu Nuthalapati, S. (2023). AI-enhanced detection and mitigation of cybersecurity threats in digital banking. *Educ. Adm. Theory Pract.,* 29(1), 357-368.
17. Banerjee, P., Roy, R., Batchu, C., & Ranjan, P. (2023). Examining the Application of Data Federation across Cloud Databases in the Financial Services Domain. *ESP Journal of Engineering & Technology Advancements (ESP JETA),* 3(1), 148-153.
18. Kommera, A. R. (2023). Empowering FinTech with Financial Services cloud. *International Journal on Recent and Innovation Trends in Computing and Communication,* 11(3), 621-625.
19. Olabanji, S. O. (2023). Advancing cloud technology security: Leveraging high-level coding languages like Python and SQL for strengthening security systems and automating top control processes. *Journal of Scientific Research and Reports,* 29(9), 42-54.
20. Oloruntoba, O., Ekundayo, T., &Aladebumoye, T. (2022). Optimizing Investments with Cloud-Based Data Mining Frameworks. *Int. Res. J. Mod. Eng. Technol. Sci,* 4(12), 2172.
21. Al-Baity, H. H. (2023). The artificial intelligence revolution in digital finance in Saudi Arabia: A comprehensive review and proposed framework. Sustainability, 15(18), 13725.
22. Parisa, S. K., Banerjee, S., & Whig, P. (2023). AI-Driven Zero Trust Security Models for Retail Cloud Infrastructure: A Next-Generation Approach. *International Journal of Sustainable Devlopment in field of IT*, 15(15).
23. Ashta, A., & Herrmann, H. (2021). Artificial intelligence and fintech: An overview of opportunities and risks for banking, investments, and microfinance. Strategic Change, 30(3), 211-222.

*********