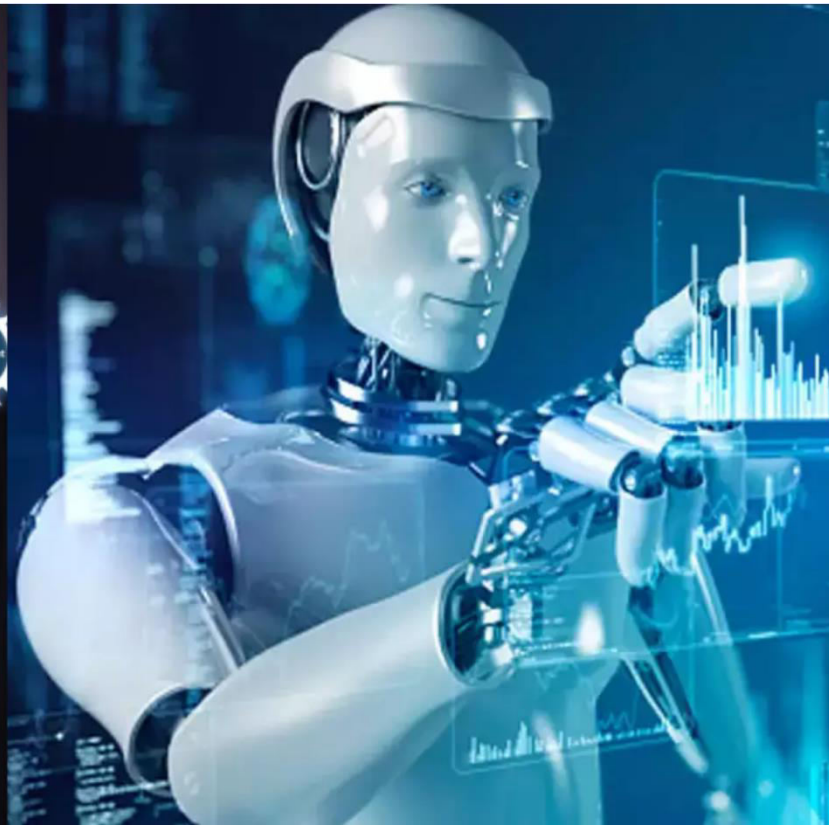
IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

ISSN(O): 2319-8753
ISSN(P): 2347-6710

International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 8, August 2025

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 8, August 2025

|DOI: 10.15680/IJIRSET.2025.1408104|

Phishing URL Detection using Graph Neural Networks and Rule-Based Models with Blockchain Integration

Ankitha L S, Sunitha G P

P.G. Student, Department of Master of Computer Applications, JNN College of Engineering, Shivamogga, India

Associate Professor, Department of Master of Computer Applications, JNN College of Engineering, Shivamogga, India

ABSTRACT: The frauds using websites and URLs are always among the extremely lethal and serious threats to online security since phishing attacks are designed to misuse the people using the application by stealing precious data. More likely to be inflexible and uninterpretable is older machine learning detectors such as blacklists and autonomous independent machine learning classifiers. This work suggest that a hybrid phishing detector, utilizing both a Graph Neural Networks (GNN) and a Rule-Based Model and backed by a blockchain history, is the optimal solution to these problems prominent for cybersecurity. GNN can identify complex graphical properties in the URLs, domains, and links among the URL nodes to achieve effective classification when a Rule-Based Model requires a heuristic rule so that the articulation of the phishing flows can be intelligible. The fusion logic merges model confidence with rules and determines final decisions that are recorded permanently on a blockchain to generate tamper-free records.

KEYWORDS: Phishing Detector, Graph Neural Networks, Rule-Based Model, Blockchain, Cybersecurity, Fusion Logic.

I. INTRODUCTION

Phishing currently ranks among the most common forms of cyber-crime and has been employed to commit fraud, steal identities and cause humongous data breaches. The whitelist approach is not effective because the classic whitelist approach would not work when the attacker continuously evolves the tactics that are being used to deliver the attack and cannot identify the newly discovered phishing websites on a zero-day basis. Machine learning models have shown promise, but are often black boxes, whose operation is difficult to comprehend and trust. This introduces an accuracy-interpretability trade-off that is among the largest problem in detecting phishing.

GNN have proven to be a useful representation of relational (relation) data such as the correlation between domains, sub domains and URLs. The phishing campaigns can be identified through structural and contextual phenomenon that cannot be associated with lexical analysis only. However, GNNs lack transparency, where security experts and database users must be able to interpret models to understand why a specific URL was flagged. Rule-based Models, though, prove to be more explainable and transparent when it comes to detection based on a combination of heuristics e.g. suspicious keywords, abnormal domain length, multiple subdomains, domain age anomalies.

To complete this task, develop a dual phishing classifier a combination of rule-driven heuristics and GNN-driven classification. Strong classification, as well as interpretation can be made on the fusion mechanism. Outside of this, to increase durability and credibility, the final decision in end-result detection and the model confidence scores and rule-based interpretation are written to a blockchain. In information systems, with this kind of an integration incentive, obstacle is discouraged and responsibility bestowed in the cybersecurity system.

II. LITERATURE SURVEY

Various number of methods have been proposed intended for phishing detection, predominantly in blockchain and URL-based attacks. Patel & Singh [1] presented a GNN framework exhibiting blockchain transaction networks, attaining 96.2% accuracy using attention-based message passing. Ao Xiong et al. [2] also used GNNs, integrating transaction attributes and multi-scale structures for Ethereum phishing detection. Similarly, Nguyen & Tran [3] demonstrated a hybrid GNN

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 8, August 2025

|DOI: 10.15680/IJIRSET.2025.1408104|

and rule-based approach, corresponding precision and computational efficiency. Zhen Chen [4] used a hybrid GNN with data augmentation but lacked in temporal sequence analysis. Rule-based models remain effective in certain contexts. Brown et al. [5] developed smart contracts with predefined heuristics for on-chain phishing prevention, while Doe & Lee [6] analysed DNS metadata using lexical and WHOIS-based features, achieving high recall. Li et al. [7] benchmarked rule-based phishing detection strategies, finding them effective against static threats but weak against adaptive attacks. Johnson et al. [8] enhances WHOIS-based detection using blockchain-backed tamper-proof records. Blockchain integration enhances auditability and resilience. Kim & Park [9] proposed a blockchain DNS system for domain verification. Garcia et al. [10] built a decentralized trust system using GNNs and rule thresholds. These studies highlight the strengths and the weaknesses of individual techniques. GNNs excel in structural pattern recognition, rule-based models provide transparency, and blockchain ensures data integrity. The proposed work in this development aims to combine all three – GNNs for intelligent detection, rule-based logic for interpretability, and blockchain for secure logging – addressing existing breaches and providing a robust solution against phishing attacks.

III. METHODOLOGY

The phishing detection system proposed works with incoming traffic that starts with a user entering a URL. Such URL is initially placed under a blockchain scan, which tells whether the URL is already processed and located in the blockchain. In case the URL is in the blockchain, the result previously stored will be called to the user, and displayed right away, thus being efficient but also trusted, because blockchain ensures that the result, known as a verdict, cannot be changed.

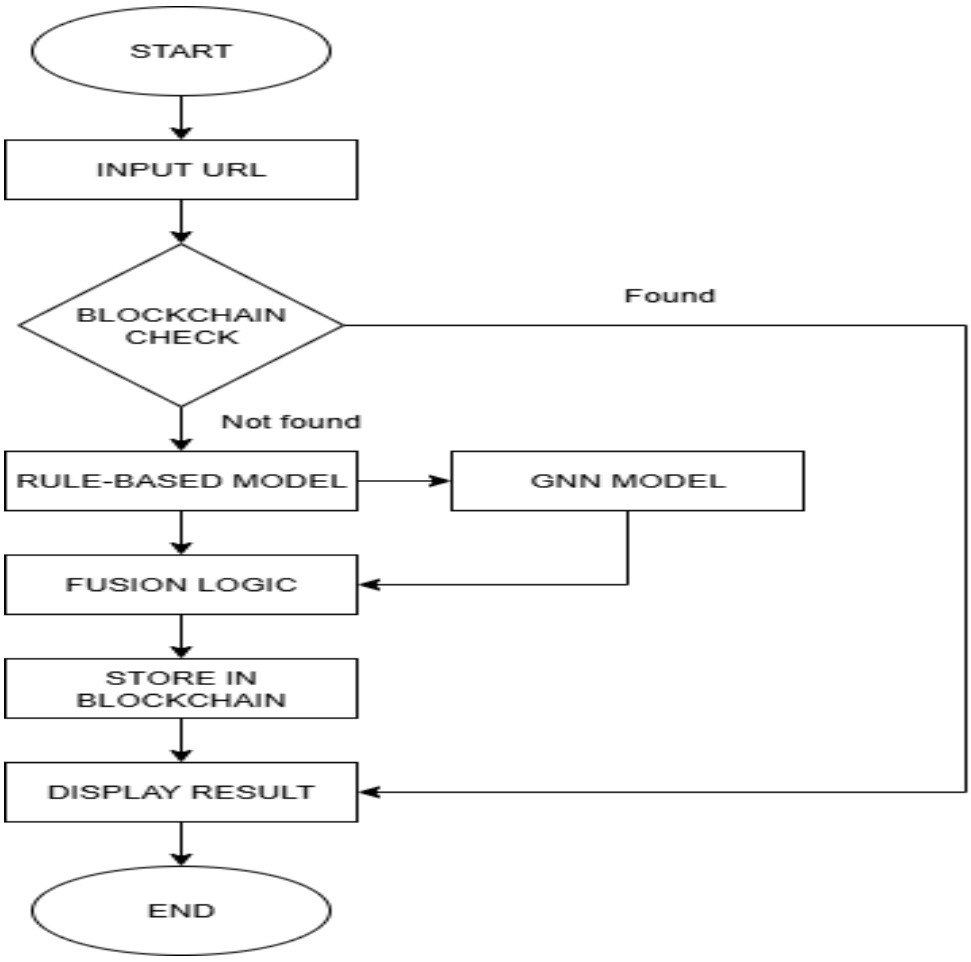


Fig 1: Architectural design

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 8, August 2025

|DOI: 10.15680/IJIRSET.2025.1408104|

As shown in Fig 1 failure to locate the URL in the blockchain will result in the system entering the detection phase, during which the URL is assessed both with the Rule-Based Model and the GNN Model. The Rule-Based Model examines the URL based on established evidence or heuristics like the inclusion of suspicious key words, unusual URL length, use of more than one subdomain or the use of IP addresses rather than domain names. This gives meaningful information as to why a URL could be malicious. Meanwhile, the GNN makes sense of the structural and relational properties of the URL by framing this as a graph of substance composed of domains, sub-domains, and hyperlink contacts.

The GNN uses the message passing and embedding generation procedure to generate a score that shows whether the URL is phishing or legitimate.

Outputs of both models are then fed into the fusion logic module which combines the merits of both approaches. Where there is consensus in both models, final-point decision is reached; where the two find no consensus, the confidence score of GNN and rule-based explanation is weighted to give more balanced decision with a warning flag in many cases raised to note uncertainties. This makes the system easily accurate, as well as transparent and understandable by security analysts and end-users.
After having made the final decision, the outcome with the label of the classification, the model score of confidence and the rule explanation are recorded in the blockchain. The step is immutable, accountable and auditable and the same URL query can be resolved immediately as opposed to hitting it again. The last step is the machine launches the result to the user in a form understandable to him and the workflow is complete.

Basically, this integrative organization makes the system to achieve a union of speed, accuracy, transparency, and security, which is especially valid in practice in the detection of phishing.

IV. EXPERIMENTAL RESULTS

The comparison of the performance of the standalone Rule-Based Model and the Graph Neural Network (GNN) outlines the advantages and shortcomings of each model. The Rule-Based Model was able to attain 67% accuracy with a high precision of 88 and a low recall of 38 achieving an F1-score of 54. This means that although the rule-based system is efficient at recognizing phishing URLs when it is configured to do so (low false positives), there is a large number of phishing attempts that it does not detect (high false negatives). Conversely, the GNN showed significantly higher overall results, being able to achieve an accuracy of 87, a precision of 82, a recall of 92, and an F1-score of 87. These findings show that the GNN is relatively successful at detecting phishing patterns, particularly due to recall or the ability to detect most of the phishing patterns. It is, however, slightly less precise than the rule-based model, which also implies that it generates more false positives. All these results demonstrate the supplementary nature of the two models: the Rule-Based Model is most accurate and interpretable, and the GNN demonstrates a high recall and high classification accuracy. As shown in Table 1.

Table 1: Standalone models performance

Model	Accuracy	Precision	Recall	F1-score
Rule-Based	67%	88%	38%	54%
GNN	87%	82%	92%	87%

By utilizing the complementing advantages of both strategies, the system's overall performance was significantly enhanced when the Rule-Based Model and GNN were integrated using the suggested fusion logic. When the Rule-Based Model and GNN were fused using the proposed fusion logic, the system demonstrated a substantial increase in the overall performance of the system due to the synergistic contributions of the two methods. While the GNN offered strong recall and accurate classification by capturing graph-based relationships among URLs and domains, the Rule-Based Model contributed its high precision and interpretability, guaranteeing that phishing detections were backed by understandable heuristic explanations. By using the GNN's capacity to detect hidden phishing patterns, the fusion method successfully decreased false negatives. By adding rule-based checks, it also decreased false positives. Consequently, the hybrid system outperformed the independent models in terms of balanced metrics, proving that the combination of deep graph learning and explainable rules results in a more reliable, accurate, and trustworthy phishing detection system.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 8, August 2025

|DOI: 10.15680/IJIRSET.2025.1408104|

V. CONCLUSION

In this study, created a hybrid phishing detection system that integrates a Rule-Based Model with GNN, backed by blockchain integration and fusion logic. While the Rule-Based Model added interpretability by offering concise heuristic explanations, the GNN showed a great ability to capture structural linkages for accurate classification. According to experimental findings, combining the two methods greatly increased accuracy and resilience when compared to standalone models, which also decreased false positives and false negatives. Furthermore, tamper-proof storage of final choices was guaranteed by blockchain integration, which improved detection process transparency and confidence. All things considered, the suggested approach offers a dependable and comprehensible phishing detection solution. Future research will concentrate on expanding the framework to incorporate phishing email detection, comprehensive webpage analysis, and adaptive learning to thwart changing attack tactics.

REFERENCES

1. Patel and Singh, “GNNs for Cybersecurity”, IEEE Transactions on Cybersecurity, 2023.
2. Ao Xiong et al., “Ethereum Phishing Detection Based on Graph Neural Networks”, in Proc. IEEE Xplore Conference, 2023.
3. Nguyen and Tran. “Hybrid AI-Rule-Based Phishing Detection”, Springer-Machine Learning & Security, 2021.
4. Zhen Chen et al., “Ethereum Phishing Scam Detection Based on Data Augmentation Method and Hybrid Graph Neural Network Model”, in Proc. ACM Conference on Cybersecurity, 2024.
5. Brown et al., “Ethereum-Based Smart Contracts for Cybersecurity”, Journal of Blockchain Technology & Applications, 2023.
6. Doe and Lee, “DNS-Based Phishing Detection”, in Proc. ACM International Conference on Security and Privacy, 2022.
7. Li et al., “Rule-Based Phishing Detection Techniques”, Springer- Cyber Treat Intelligence Review, 2020.
8. Johnson et al., “WHOIS-Based Phishing Detection”, ACM Transactions on Information and System Security, 2020.
9. Kim and Park, “Blockchain DNS Security”, Elsevier-Computer & Security, 2022.
10. Garcia et al., “Blockchain-Based Trust and Reputation System”, Springer-Cybersecurity Journal, 2022.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details