Blockchain-Aided Spoofed URLs Detection: A GNN and Rule-Based Fusion

Sunitha G P[1], Ankita L S[2], Santhosh S G[3]
[1]*Assoc. Prof., JNNCE, Shimoga*
[2]*PG Student, Dept. of MCA, JNNCE, Shimoga*
[3]*Assoc. Prof., Dept. of MCA, JNNCE, Shimoga*

doi.org/10.64643/IJIRTV12I4-184568-459

***Abstract*—Phishing attacks have been a major threat in cybersecurity, plays with human trust by deceiving users to interact with fraudulent websites and imposes on technical vulnerabilities. Traditional detecting systems, such as signature-based rule- based methods, This often fail to identify original and obfuscated phishing URLs. This paper presents an intelligent and secure phishing detection system which combines the capabilities of Graph Neural Networks (GNNs) and rule-based feature analysis, integrated with the blockchain technology for fixed logging. The GNN model analyses the structural relationships within the URL using a graph-based representation, on the other hand the rule-based module examines key heuristic indicators which includes the presence of IP addresses, URL length, and HTTPs usage. Each prediction of GNN is logged into a private blockchain using a proof-of-work consensus to ensure transparency, traceability and for tamper-resistant. Additionally, disagreement cases between models are highlighted with rule- based explanations and blockchain traceability.**

***Index Terms*—Phishing Detection, GNN, Rule-Based Model, Blockchain, URL Classification, Cybersecurity**

I. INTRODUCTION

Phishing attacks are a growing cybersecurity concern, aiming to steal sensitive data such as login credentials and banking details by impersonating trusted sources. These attacks are commonly delivered through malicious URLs in emails, messages, or fake websites, targeting both individuals and organizations, leading to financial losses and data breaches. Traditional phishing detection relies on blacklist-based systems, which are fast but ineffective against new or short-lived phishing sites. To improve detection, rule-based methods analyze URL structures—like unusual characters, IP usage, or absence of HTTPS—but lack adaptability to evolving threats and may cause false alarms.

Machine learning models, counting SVMs and Decision Trees, offer better generalization but treat URLs as flat feature vectors, missing important structural patterns. Graph Neural Networks (GNNs) address this by modelling URLs as graphs, capturing complex relationships between URL components. However, GNNs act as black-boxes, offering little interpretability.

To overcome this, the proposed system combines GNN-based detection with a parallel rule-based module that extracts explainable features. A fusion logic merges results from both models, highlights disagreements, and provides justifications to improve trust. To ensure decision integrity, predictions are logged into a private blockchain. This immutable ledger stores URL, label, confidence, and timestamp, providing transparency for audit and compliance.

The system is deployed via a Flask web interface, offering real-time classification, confidence scores, and rule-based explanations. By integrating GNNs for intelligence, rule-based logic for clarity, and blockchain for security, the system provides a robust, transparent, and adaptive solution to phishing detection.

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 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] analyzed 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 system integrates rule-based heuristics, graph neural networks (GNNs), and blockchain technology which forms a robust phishing URL detection framework. The methodology is planned in a way that it not only predicts phishing URLs but also safeguard that all the results are stored in tamper-proof blockchain for forthcoming verifications.

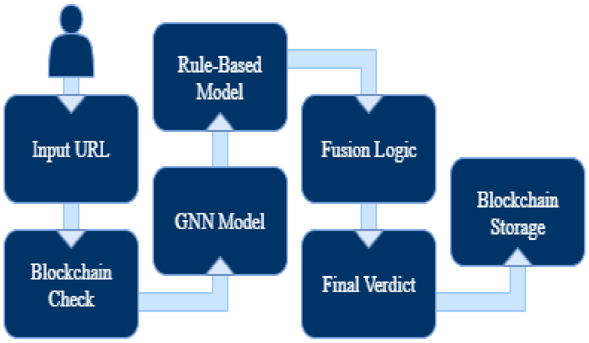


Fig 1: Architectural Design

The work flow be broken down into the stages follows:

3.1 Input URL

The process begins in very first stage where user enters a URL and submits for verification. At this point, the system performs an initial blockchain lookup. This ensures that if in the past the URL has already been analyzed, the previously stored result can be retrieved instantly. This mechanism speeds up the detection process by reducing the computational overhead. If the URL is not found in the blockchain, the system proceeds for fresh detection using rule- based and GNN models.

3.2 Rule-Based Feature Extraction and detection

The rule-based model acts as first layer of defense. It relies on carefully designed heuristics commonly associated features with phishing URLs. Examples include:

- Length of the URL.
- Presence of the special characters whereas @, //, -, or may have excessive subdomains.
- The URL uses a secure protocol (https) or not.

Each of these features is assigned some weight based on its importance, and a score is computed. If the score surpasses a threshold, the URL is flagged as phishing. Although this method is lightweight and efficient, it may may miss sophisticated phishing attempts.

3.3 Graph Construction and GNN-Based Detection

The Graph Neutral Network (GNN) forms a backbone of advanced detection. The GNN reflect relationships among URLs, as an alternative of analyzing them in isolation. There by, URLs are characterized as nodes in graph, and edges are constructed based on their resemblance. The graph structure permits the GNN to capture hidden patterns. For example, phishing URLs every so often share domain structures, keyword patterns, or IP-based hosting resemblances. Through message passing, the GNN absorbs embeddings for each node by collecting information from its neighbors.

Finally, the learned embeddings are passed through a classifier that forecasts whether a URL is legitimate or phishing. This model is highly effective at detection of sophisticated and earlier unseen phishing attempts.

3.4 Fusion Logic

Subsequently neither rule-based methods nor GNNs are perfect on their own, the system introduces a fusion mechanism to combine their strengths.

- If both models agree, the decision is straightforward.

If one predicts phishing while the other predicts legitimate, the system errors on the side of caution and classifies the URL as phishing. This conservative approach minimizes false negatives, which are further more dangerous in phishing detection.

3 The confidence score is calculated as the average of the individual model confidences, providing users with a measure of reliability.

i. Rule-Based Model Confidence

4 The rule-base model examines a set of N handcrafted phishing detection rules. If the URL triggers n of these rules, then the confidence score is:

$$Conf(R(u)) = \frac{n}{N} \text{--------} eq\ (1)$$

ii. GNN Model Confidence

The GNN produces probabilistic output via the SoftMax activation function:

$$G(u) = [p_{(legit)}, p_{(phish)}] \text{-------} eq\ (2)$$

The phishing confidence is taken as:

$$Conf(G(u)) = p_{(phish)} \text{-----------} eq\ (3)$$

iii. Final Prediction Rule

The decision-making follows a conservative strategy:

$$Final(u) = \begin{cases} phish, if R(u) = phish \lor G(u) = phish \\ legit, if R(u) = legit \land G(u) = legit \end{cases}$$

$$\text{- --} eq\ (4)$$

Thus:

- If both models agree, takes that decision.
- If both models disagree, marks as *phishing*.

IV MATHEMATICAL MODEL

The process of identifying URLs as phishing or legitimate is described in this section.

URLs to be represented as set of Input URLs, whereas:

a. *U = {u1, u2, u3, ..., un}* be the input URLs set.

b. Each *URL ui* is represented as a feature vector *F(ui) Є* R^d.

c. *ui* = i^{th} URL node in the dataset

d. *F(ui)* = vector representation of the URL

e. R^d = *d*-dimensional real vector space, where *d* is the embedding dimension (e.g., 16, 32, 64)

This indicates that each URL is represented as a vector of real numbers that capture semantic and structural features useful for classification.

4.1 Rule-Based model

Handcrafted heuristic functions are defined, where these heuristic functions apply rule-based features to assign a phishing probability score to a given URL:

$$R(u_i) \begin{cases} 1, if \sum_{j=1}^{m} w_j f_j(u_i) > \theta_r \\ 0, \quad otherwise \end{cases} \text{-----------} \mathrm{eq}\ (5)$$

Where:

Where $p(legit) + p(phish) = 1$

- *fj(ui)* = extract feature *j* (e.g., length of URL, presence of *"@"*, HTTPS usage, suspicious keywords).
- *wj*= weight assigned to feature *j*.
- *Ɵr* = threshold that is defined set a target to distinguish between Phishing or Legitimate.
- Output: *R(ui) Є {0 (Legitimate), 1 (Phishing)}.*

For example, if the length of the URL exceeds 75 characters, classifies it as phishing.

Let:

- Threshold *Ɵr* =75 characters
- U = https://www.google.com/accounts/login
- Length = 35 < 75
- *R(ui)*= 0, which classified as Legitimate

4.2 Graph Neural Network (GNN) Model

GNN model is a graph-based model. Int the graph construction, each URL is represented as a node connected to feature-based entities. This structure allows the GNN to capture both individual URL features and their contextual relationships for effective phishing detection.

Let construct a graph *G = (V, E),* where:

- *V = {v1, v2, ..., vn}* represents URLs as nodes.
- *E = V × V* represents similarity relations between URLs.

The GNN updates node embeddings using message passing:

$$h_v^{(k)} = \sigma\left(\sum_{u \in N(v)} \frac{1}{c_{(vu)}} W^{(k)} h_u^{(k-1)} + W_0^{(k)} h_v^{(k-1)}\right)$$

$$\text{-----} eq\ (6)$$

Where:

- $h_v^{(k)}$ = embedding of node *v* at layer *k*.
- *N(v)* = neighbourhood of node *v*.
- $W^{(k)}$, $W^{(k)}$ =$_o$learnable weight matrices.
- C_{vu} = normalization factor.
- σ = non-linear activation

The final classification layer:

$$G(u_i) = Softmax(Wh^{(k)}_{ui}) \text{ ---------} eq\ (7)$$

Output: *G(ui) € {0,1}*.

The softmax method is used to convert raw output scores into likelihood of probabilities that sums up to

$e^{1.0}$= 2.71.

- Sum = 7.39 + 2.71 = 10.10
- Legitimate = 0.26
- Phishing = 0.73

4.3 Fusion of Results The hybrid system fuses predictions:

$$F(u_i) = \{ \begin{matrix} 1, if\ (R(u_i) = 1) \vee (G(u_i) = 1) \\ 0, otherwise \end{matrix} \text{ --------------------} eq\ (4)$$

Confidence Score, where it is the probability output from GNN and Rule-Based model indicating how strongly they believe a URL is Phishing or Legitimate:

$$C(u_i) = \frac{Conf(R(u_i)) + Conf(G(u_i))}{2} \text{-------------} eq\ (5)$$

Where *Conf ()* is confidence probability of each model.

4.4 Blockchain Storage

Each final decision F(ui) is stored as a transaction:

$B_t = H(F(u_i) \| C(u_i) \| u_i \| Prev\_Hash)$ -- eq (6)

- H () = cryptographic hash function.

Where:

- ---------*eq (6)*

- Prev_Hash = hash of previous block.
- Ensure immutability and tamper-proof record.

V EXPERIMENTAL RESULTS

5.1. Dataset

The system evaluated using URL dataset containing 2,000 URLs, which comprises both legitimate and phishing websites. The dataset has been distributed into 80% for training and 20% for testing for the GNN model. The rule-based model was tested on the complete dataset using handcrafted detection features.

1. For example, suppose the model output the score as [2.0,1.0] for two classes as Phishing and Legitimate.

- Computes exponential: $e^{2.0}$ = 7.39,

id	Previous hash	timestamp	data	nonce	hash
9	0000ee6c3235c3e55af62f44 07b92ed8d92fb5ae54c796d 7ddec635849a3b046	1750099645	{"confidence": 5, "feature_hash": "2c19d5808490e5cc", "prediction": "LEGITIMATE", "timestamp": "2025-06-17T00:17:25.963251", "url": "https://www.youtube.com/"}	44153	0000e687d369c15a 9d05be2db9ac56be 9ea964edcda29b1e ad77823aa7151835

Table 1: Block chain storage

5.2. Evaluation Metrics

The performance of the models was measured using the following metrics:

Table 2: Metrics used for performance evaluation

Accuracy	$\frac{TP + TN}{TP + TN + FP + FN}$ .... $eq(7)$
Precision	$\frac{TP}{TP + FP}$ ...... $eq\ (8)$
Recall	$\frac{TP}{TP + FN}$ ... . $eq(9)$
F1	$\frac{2 \cdot Precision \cdot Recall}{Precision + Recall}$ ... ... $eq(10)$

Where *TP, TN, FP,* and *FN* signify true positives, true negatives, false positives and false negatives respectively.

5.3. Model Performance

The performance comparison between the models

Model	Accuracy	Precision	Recall	F1-score
Rule-Based	66%	89%	39%	54.3%
GNN	87.5%	83%	92.7%	87.8%

Fig 2: Performance of Individual Models

The experimental results highlight the complementary strengths of the two models. The rule-based model exhibits high precision (89%), indicating reliability in detecting phishing when flagged, but suffer from low recall (39%), missing a significant number of phishing URLs. Conversely, the GNN model achieves high recall (92.7%), effectively capturing most phishing attempts, but at the cost of reduced precision (83%).

To address these trade-offs, the proposed fusion logic integrates both models by averaging their confidence scores. This approach leverages the precision advantage to the rule-based model and the recall advantage of the GNN model, resulting in a more robust final decision. Consequently, the fusion method yields superior overall performance, achieving higher accuracy and F1-score in comparison to the individual models.

5.4. Blockchain Integration

The blockchain component was estimated for its capability to store and retrieve previously detected results.

5.5. Analysis

Analyzing the descriptive ability of the GNN model, visualized the node embeddings using t-SNE (t-distributed Stochastic Neighbor Embedding). Fig 3 illustrates the distribution of phishing and legitimate URLs in the GNN output space. The visualization clearly displays that phishing URLs (red points) and legitimate URLs (green points) form distinct clusters, although some overlap regions still exist. This demonstrates that the GNN effectively learns meaningful representations of URL structures and relationships, which are critical for accurate classification. The clusters confirm the model's high recall (92.7%), as most phishing samples are grouped together. However, the small overlapping areas explain why the GNN occasionally misclassifies certain legitimate sites, aligning with the precision (83%) observed in quantitative results.

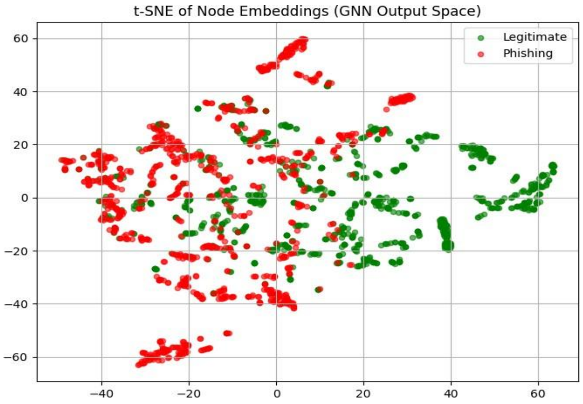
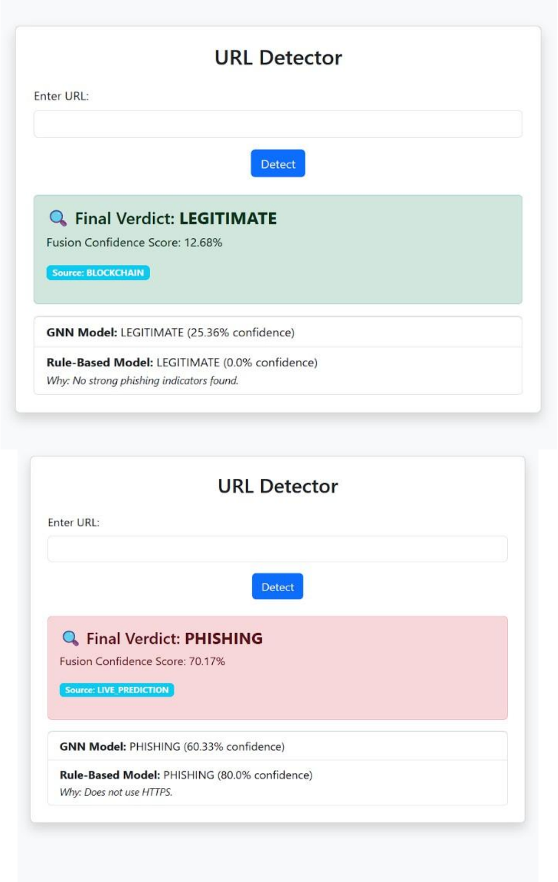


Fig 3: t-SNE output

By combining this GNN-based representation learning with the rule-based model through fusion, we reduce misclassification in the overlapping region, improving overall system robustness.



URL Detector

Enter URL:

Final Verdict: **LEGITIMATE**

Fusion Confidence Score: 12.68%

Source: BLOCKCHAIN

GNN Model: LEGITIMATE (25.36% confidence)

Rule-Based Model: LEGITIMATE (0.0% confidence)

Why: No strong phishing indicators found.

URL Detector

Enter URL:

Final Verdict: **PHISHING**

Fusion Confidence Score: 70.17%

Source: LIVE_PREDICTION

GNN Model: PHISHING (60.33% confidence)

Rule-Based Model: PHISHING (80.0% confidence)

Why: Does not use HTTPS.

Fig 3: Website Outlook

By combining this GNN-based representation learning with the rule-based model through fusion, we reduce misclassification in the overlapping region, improving overall system robustness.

VI CONCLUSION

The phishing detection system integrates a Graph Neural Network (GNN), a rule-based model, and blockchain to achieve both accuracy and security. The rule-based model guarantees coverage against zero-day assaults, but at the expense of more false positives, whereas the GNN efficiently detects intricate phishing patterns with good recall. By utilizing fusion logic, the system strikes a balance between safety and accuracy, lowering false alarms while preserving robust detecting capabilities. Blockchain also makes it possible to save previous detection results in a way that cannot be altered, which promotes trust and reuse. In practical applications, these elements work together to create a dependable, transparent, and strong defense against phishing attacks.

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.