LEX LOCALIS-JOURNAL OF LOCAL SELF-GOVERNMENT
ISSN:1581-5374 E-ISSN:1855-363X
VOL. 24, NO. 3 (2026)

GOVERNANCE-AWARE PHISHGUARD++: A ZKP-DID AND GRAPH NEURAL NETWORK FRAMEWORK FOR ACCOUNTABLE CYBERSECURITY IN LOCAL DIGITAL PUBLIC SERVICES

K. Ramya[1*], S. Anbukaruppusamy[2]

[1*]Department of Instrumentation and Control Engineering (ICE), Anna University, Chennai – 600025, India
[2]Department of Electronics and Communication Engineering, Excel Engineering College, Komarapalayam – 637303, India

*Corresponding Author: K. Ramya , Email: rythmramya@gmail.com

Abstract: Digital technologies are increasingly used in the delivery of local public services, which makes it important to have strong authentication, privacy, and accountable access control on digital governance platforms. In this study, an authentication solution named Governance-Aware PhishGuard++ from the Web3 space is proposed that combines Zero Knowledge Proofs (ZKPs), Decentralized Identifiers (DIDs), and Graph Neural Network (GNN) based threat intelligence to augment the identity verification process, safeguard sensitive information, and enhance the capability of phishing threat detection. The principal contribution is called a Governance-Aware Risk-Adaptive Access Control (GRAAC) mechanism, which combines identity verification, dynamic phishing-risk estimation, institutional authorisation policies, and accountable access enforcement. The proposed extension differs from the original framework in that it also considers factors such as the probability of a threat, institutional permission, resource sensitivity, and governance requirements in its evaluation of access requests rather than just the mutual authentication and real-time threat detection in the original framework. Only a policy-constrained decision mechanism will decide whether to allow or disallow requests as well as whether to escalate to authorised review. Enforcement and privacy-conscious audit records via smart-contracts or the application level support traceability without necessarily revealing sensitive identity information. It will be tested using the available phishing data sources and synthetic municipal-services phishing scenarios. The experiments will compare accuracy of phishing detection, unauthorised access acceptance, denial of legitimate access requests; policy compliance, latency in decision making and computation overhead; and completeness of auditing. The proposed mechanism will be evaluated by means of baseline comparison, ablation studies, and sensitivity analysis. The research aims to combine explicit institutional governance policies with privacy-preserving authentication and graph-based threat intelligence to lay a foundation for measurable and accountable cybersecurity for local digital public services.

Keywords: PhishGuard++; Zero-Knowledge Proofs; Decentralized Identifiers; Graph Neural Networks; Risk-Adaptive Access Control; Local Digital Governance.

1. Introduction

Local public services are now digital and are much more accessible and efficient, but there are new challenges in the cybersecurity area, such as phishing, identity theft, and unauthorized entry. Solely traditional security would be enough against high-level threats and to uphold privacy. This study, in turn, proposes Governance-Aware PhishGuard++, an advanced cybersecurity framework that combines Zero-Knowledge Proofs (ZKPs), Decentralized Identifiers (DIDs), and Graph Neural Networks (GNNs). The ZKP-DID mechanism enables users to verify credentials without revealing sensitive information, and the GNN can detect suspicious activities and potential phishing threats based on the analysis of interactions.

In addition, it features a Governance-Aware Risk-Adaptive Access Control (GRAAC) system that can control access based on identity verification and threat assessment, according to institutional policy and user role. It is concerned about auditability, privacy, and compliance, aims to make phishing detection and access control more reliable, and protects citizen data. There are various metrics to measure the

effectiveness of Governance-Aware PhishGuard++, and ultimately it can drive towards a privacy-preserving and governance-oriented local digital public services security.

2. Related Work

The literature started with traditional techniques for phishing detection, then moved towards ensemble learning and deep neural network approaches. In, Yang et al. (2021) studied the use of a deep convolutional network and a random forest for phishing website detection, adding an extra layer of relevance to the use of learned representations, along with conventional classification. Ozcan et al. (2021) studied a phishing URL detection model based on a hybrid system of a deep neural network and long short-term memory, which showed that the hybrid model can be used to detect various features of malicious URLs. Esposito et al. (2021) explored blockchain-based security and authentication/authorization for smart-city applications, linking blockchain security mechanisms to smart city operational requirements of interconnected digital services. Regarding access rights management in smart cities, Sabrina and Jang-Jaccard (2021) studied access control through smart cities' blockchain technology based on entitlement. These studies are complementary and collectively form two research strands: detection in the online information space and control of access to online resources. But phishing classification and access authorisation are answering different security questions; moreover, by achieving high classification accuracy alone, it is not possible to ensure that an access request is compatible with the institutional rules. The separation is the drive for considering detection and authorisation as related but separate functions.

Further efforts have studied phishing detection methods using various fusions of ML, DL, and security analytics. To illustrate the need for testing the detection methods, Sánchez-Paniagua et al. (2022) analyzed the detection of phishing URLs in a real case scenario of login URLs. The systematic review of Catal et al. investigated the use of deep learning in the patterns of phishing and the methodological issues in it. Bagga et al. (2022) mentioned access control in IoT applications using blockchain and the paradigm of blockchain for access control in a distributed environment. Since the decentralised architecture is based on the assumption of 'Self-Sovereign ID Verification', adding fine-grained access control and privacy protection as proposed by Ma et al. (2022) is difficult for identity verification. Asif et al. (2022) discussed blockchain applications to authentication and trust management for smart cities, and Singh et al. (2022) studied several access-control schemes using blockchain techniques that aim to defend IoT systems against cyberattacks. Prolonged studies collectively demonstrate that digital security with respect to detection, authentication, permission management, and trust means nothing that can be ignored. The performance for each of the functions can be measured by different sets of data and metrics; however, it isn't simple to determine the effectiveness of these elements when they are all working together as a single workflow for accessing public services.

Maintaining privacy and, at the same time, adding user validation of the user has become a primary research topic, and the aim is to minimize the information that must be given to a service before it can be used for the user to maintain privacy. Di Francesco Maesa et al. (2023) explored self-sovereign identity and access control using blockchain technology, with the ability to prove the absence of a private attribute, using the ZK approach. Alangot et al. (2023) studied decentralised ID authentication focusing particularly on auditability and privacy, pointing out the need for accountability of authentication processes, but avoiding unnecessary privacy loss for users. Further Chi et al. (2023) investigations into cryptographic systems that can help in proving statements without revealing all of the data include privacy-preserving zero-knowledge proofs (ZKPs) specifically for blockchain applications. Some contexts in which these approaches may apply are public-service platforms during which the person may be required to put up proof for eligibility, ID, or permission. But privacy-preserving identity verification cannot alone decide whether there is safety with respect to the changing threat situation, or whether permission is granted based on the institution's access policies. Therefore, a complete security workflow should, besides verifying is requesting access to certain data, include a measurement of threat and apply restrictions according to the organisation's policy.

A second important area of detecting malicious behaviour in interconnected digital environments is graph-based security analytics. Xiong et al. (2023) came up with a graph-based feature learning approach using graph neural networks to detect phishing activity on Ethereum to identify suspicious accounts based on relationships with transactions. Safi et al. (2023) have presented an overview of

phishing website detection techniques and have reported about the ongoing evolution of techniques to detect fraudulent websites. To tackle challenges in blockchain phishing detection, some research interest has included the use of graph representations with learning strategies, as done by Chen et al. (2023) for phishing scam detection in Ethereum through data augmentation and a hybrid graph neural network. To further demonstrate the use of relational information to recognize suspicious entities, Mohan et al. (2023) studied the detection of phishing accounts on Ethereum using a graph neural network. The authors of Yan et al. (2023) summarised the current state of the art of decentralised identity management through blockchain and pointed out the research directions in this emerging field. These studies corroborate the relevance of graph learning for modelling relationships between entities, and that of decentralised identity for digital trust management. But phishing detection, institutional authorisation and ID management are separate activities. A decision mechanism is necessary when they are used in combination to map threat assessments to access outcomes whilst maintaining policy imperatives and accountability.

As well, recent studies have addressed cybersecurity in smart cities and decentralised technologies with threat-aware authentications. Smart cities also have security concerns due to the interconnected services provided by the cities, and Houichi et al. (2023) explored cybersecurity in smart cities and looked at approaches based on intrusion detection. Ramya and Karuppusamy (2026) designed and developed a privacy-preserving Web3 authentication framework called PhishGuard++. They have been the ones who directly contributed to the continuity study developed in this paper. The extension developed here is centred on governance-aware risk-adaptive access control, taking into account both identity verification and threat scores calculated from the graph, as well as institutional permissions and resource sensitivity, and compulsory governance regulations. The proposed extension will not only grant access but will accommodate actions of permit/deny/escalate based on explicit policy conditions and a low threat score. It also takes into account the following aspects in its evaluation framework: Unauthorised acceptance, legitimate-request denial, policy compliance, computational overhead, audit completeness, and robustness. This guidance applies to privacy-focused Web3 security as well as to the accountability standards of local digital public services and serves as a foundation to explore the real-world considerations of security, privacy and service accessibility.

3. Methodology

The method proposes and analyses the Governance-Aware PhishGuard++ which is an extension of the original privacy-preserving mutual-authentication scheme. The proposed research preserves the ZKP-DID-based identity verification and GNN-based threat intelligence and adds Governance-Aware Risk-Adaptive Access Control (GRAAC) to relate threat estimates with institutional authorisation policies. Exploring the research design will find that the phishing-risk estimation, identity verification, policy compliance, and access enforcement are distinctly separated, so that the contribution of each can be evaluated separately. The publicly available phishing datasets can be used to develop threat models, and the published synthetic, municipal-service scenarios can be used to test governance policies. It involves comparative baseline, component ablation, security evaluation, and computational-performance analysis in the experimental protocol. This separation allows the study to calculate the results of governance-aware access control to measure whether it enhances the effectiveness of the institutional policies, with accuracy of phishing detection as a necessary but insufficient indicator of overall system effectiveness.

3.1. Dataset Collection and Experimental Data Preparation

This research specifically focuses on the data that were related to phishing websites, and general website activities. Moreover, it will more concentration on the users interactions with the online services that particularly representing on the institutional governance. The used dataset contains 11.055 records with 30 features for the primary experiments. Usually, this can be used for classification; we need to check whether it can be adopted for Graph Neural Networks (GNNs). However, the UCI dataset also function as a additional evaluation. Also, the supplementry threat information can be analyse from the URL reports. It can be demonstrated in the belwo figure (Fig 1).

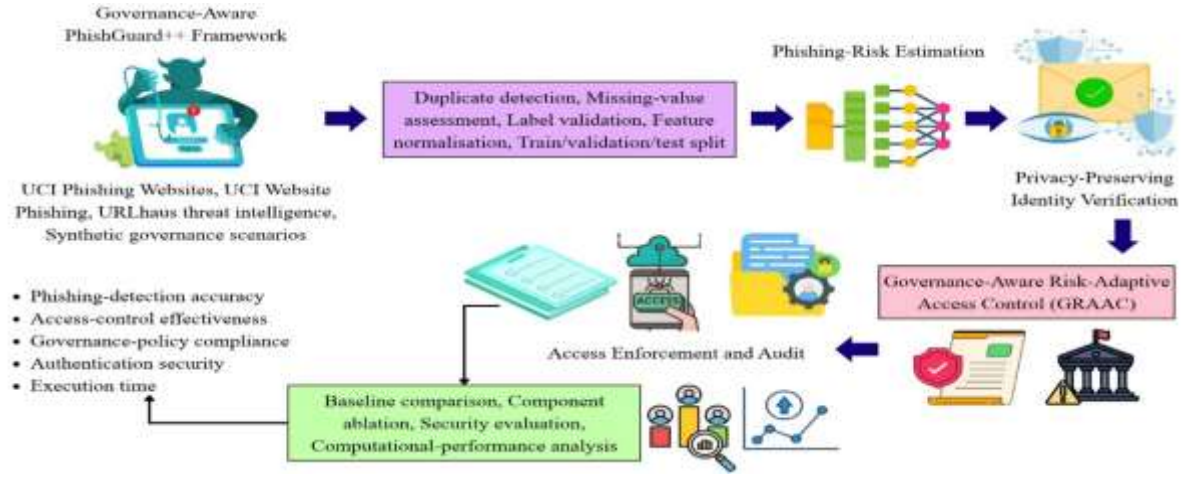


Fig 1. Dataset Collection and Experimental Data Preparation Workflow

The collection of dataset were undertaken to a peculiar preparation that include the qulaity check, lable analysation, feature normalization. This process also documented the graph components and the dataset information of training, validationa nd testing sets. This can be prevent form the data leackaging while the process undergoing. After preparing the datset, it will group by the time order and sources such as number of datset, fake data, validation of data and the distribution of the data. This proposed framework can easily identify the malicious data with the URL. This approach can resolve the phishing datset detection while the institutional governance in the control.

3.2. System Architecture and Threat Modelling

3.2.1. System Architecture

The proposed system is built up with five components which are closely interrelated: a ZKP-DID authentication module, a GNN threat-intelligence module, the GRAAC policy engine, an access-enforcement layer and an accountability audit module. The authentication module provides authentication to the necessary identity claims and service credentials. The threat-intelligence module indicates the likelihood that the current interaction is related to phishing or malicious activity. The GRAAC engine takes into account this estimate and institutional permissions, service-specific requirements, and resource sensitivity. Lastly, the enforcement layer is responsible for enforcing the resulting access decision, and an audit module keeps track of necessary minimum information for authorised review (Fig 2).

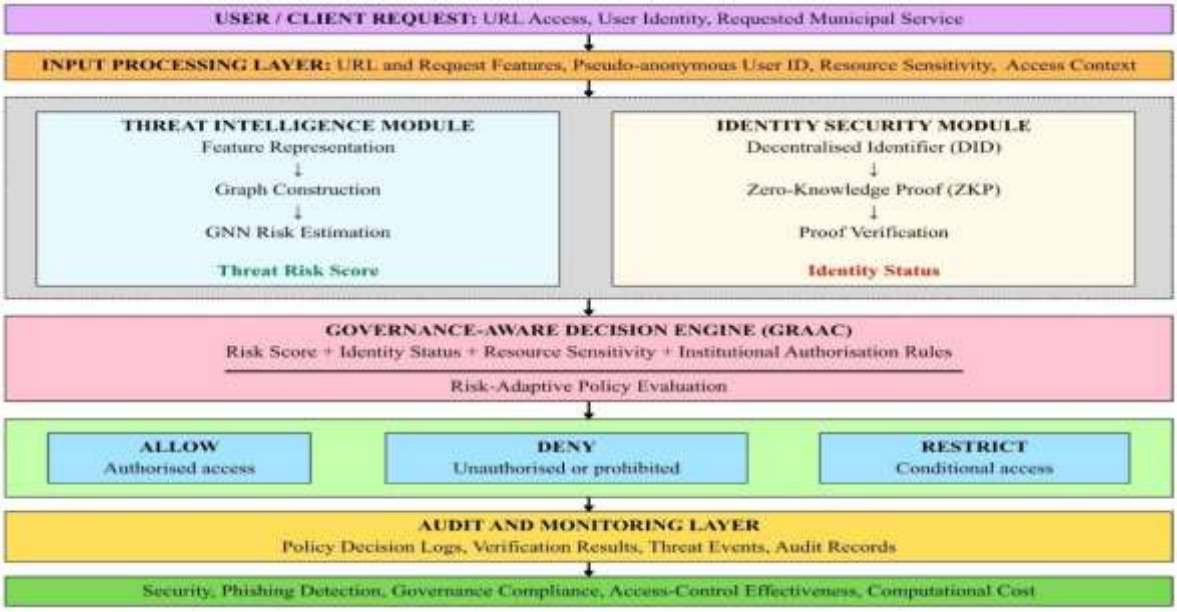


Fig 2. System Architecture of the Proposed Governance-Aware PhishGuard++ Framework.

While DIDs offer the means to decentralise identifiers and related verification, ZKPs enable a verifier to verify certain claims without always getting the private attributes. The method of implementing the DID, the credentials that are used, the proof system, the cryptographic assumptions, and the key-revocation procedure must be specified. A DID and/or a proof is not sufficient to prove that a requested action is institutionally authorised. The proposed architecture thus considers the identity verification and policy authorisation as separate actions.

3.2.2. Threat Model and Trust Assumptions

The threat model takes into account the possibility of phishing sites, spoofed public-service sites, stolen credentials, unusual access patterns, malicious nodes in the interaction graph, and potential interactions between validators (if they are part of the deployment). Threat actors might try to spoof a legitimate service, use a stolen username and password, or alter graph connections, or get contents outside of their permissions. The design relies on the correct implementation of the chosen cryptographic primitives, and the security assumptions made about the relevant proof-verification and key-management procedures.

The GNN is not viewed as a trusted and paternalistic source of identity and authorisation, but rather as a data-imperfect predictor component. Its estimates may be affected by incomplete information, spatial/temporal shift in distribution, malicious manipulation, and misclassification. This means that a low threat score won't be able to override a failure to pass identity verification or pass through an automatic policy violation. High risk / uncertain requests will be addressed via the predetermined denial/escalation rules, and will include a specific authorisation process and an auditable justification if it is denied by administration.

3.3. Experimental Design and Comparative Evaluation Protocol

The experimental design compares the proposed GRAAC framework with three baselines: conventional role-The experimental design is comprised of the proposed GRAAC framework, three baselines (conventional role-based access control (RBAC), risk-based access control with the proposed governance extension turned off, and the original PhishGuard++ configuration, and a baseline with no access control. Where technically possible, each baseline will be built-in with well-defined capabilities and equivalent access scenarios. This comparison will show the performance of the phishing-classification component and the performance of the entire access-control workflow.

Precision, recall, F1 score, ROC-AUC, along with false positive and false negative rates, will be used to evaluate the phishing classification. The evaluation of access control will deal with unauthorised access acceptances, legitimate-request denials, policy-compliance rate and decision consistency. Latency of authentication, end-to-end decision latency, computational overhead, and smart-contract gas consumption (when using blockchain enforcement) will be factors to be evaluated in the performance evaluation. The completeness of the audit will be assessed using a pre-determined checklist of Decision Records, Policy references, Review information, etc.

Ablation experiments will make use of which removes the sensitivity of the GNN risk input, governance-policy layer, or adaptive decision mechanism separately to estimate its contribution. In the event of stochastic training, the experiments will be repeated with reporting seeds and uncertainty will be presented using appropriate confidence intervals. The experimental design will be paired for statistical comparisons, which will be based on the distribution of the outcome measures. To reduce evaluation bias, test data will be separated from the threshold selection and hyperparameter tuning, and from the learning of policy.

3.4 Proposed Technique: Governance-Aware Risk-Adaptive Access Control (GRAAC)

The main methodological extension of PhishGuard++ is the proposed GRAAC mechanism. It is designed to map the findings from identity-verification processes and threat estimates from GNN to access decisions that meet institutional governance requirements. In addition to the estimated interaction risk, GRAAC considers mandatory policy constraints, which are different from traditional risk-based access control. It has the following four operations: threat-risk assessment, governance-

policy assessment, constrained access-decision making, and accountable enforcement. The mechanism does not entail that a probabilistic prediction acts as an identity, intent or legal authority; each of these is addressed by its respective verification or policy process (Fig 3).

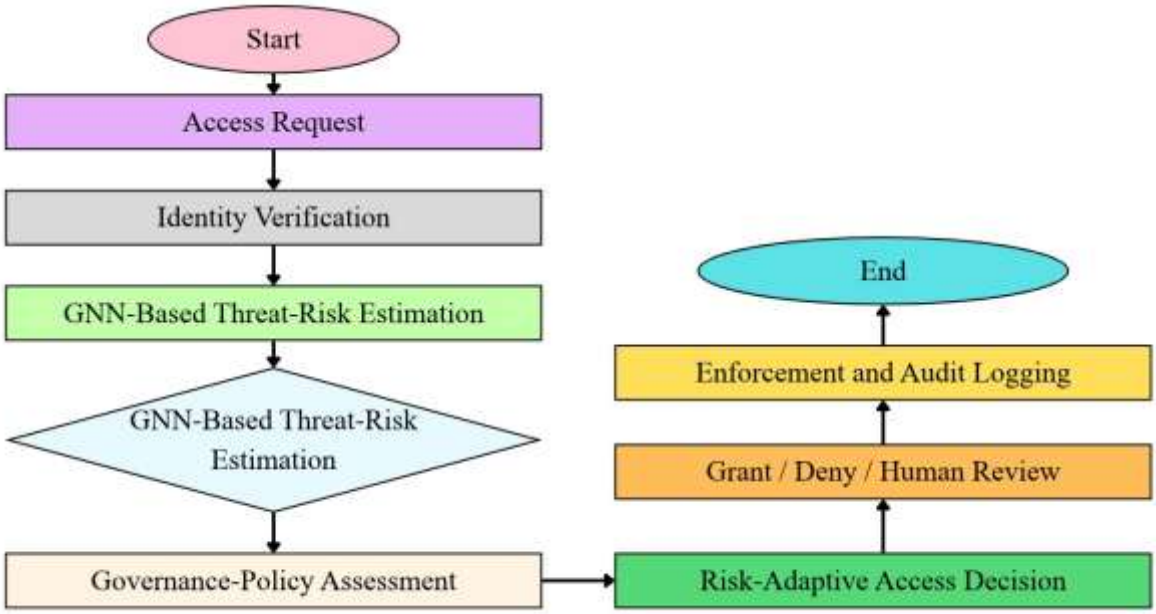


Fig 3. Governance-Aware Risk-Adaptive Access Control (GRAAC) Framework.

3.4.1 Graph-Based Threat-Risk Estimation

An interaction environment is represented as a graph G = (V, E), where V is a set of those elements involved in the interactions and records of interactions, and E is a set of observed interactions between elements in V. A feature vector h_i is associated with each node v_i and created by using information available at the time the access request is made. To learn a representation of potentially suspicious interactions, a GNN is used to combine the properties of neighbouring nodes. The estimated threat probability is:

$$r_i = \sigma(W_o \cdot h_i^{(L)} + b_o) \quad (1)$$

Where h_i-^{L} is the learned representation of the last layer of the GNN, and W_o and b_o are the parameters of the output layer, and $\sigma(\cdot)$ is the sigmoid function. The resulting $r_i \in [0,1]$ will be the model's estimate of the probability of an interaction being associated to a phishing attack or malicious interaction according to its training definition. The model will be trained with labelled data and validated by held out observations in which the selection of thresholds will be done on validation data only.

3.4.2 Governance-Policy Compliance Assessment

On every request, the governance engine checks if the requested operation will or will not comply with the institutionally applicable rules. Define q_{ik} | i = 1, 2, ..., n | k = 1, 2, ..., m | as the score of request i for criterion k, with w_k as its weight, and both score and weight are in the interval [0, 1]. The aggregate governance-compliance score is the sum of:

$$g_i = \Sigma_{k=1}^{K} w_k \cdot q_{ik}, \quad \Sigma_{k=1}^{K} w_k = 1 \quad (2)$$

Here, K is the number of the considered criteria, and g_i (for i = 1, ..., K) is the weighted compliance score of the respective criteria and is in the range [0, 1]. Role authorisation, eligibility of the services, sensitivity of the resources, and contextual access factors can be considered. Before testing, the weights should be determined via a documented procedure for designing the policy.

The overall score does not take the place of required authorisations. Missing mandatory permissions, revoked credentials, and prohibited operations will also be considered hard policy violations, as will invalid identity proofs. Even if they achieve high scores on other governance criteria, such requests can, however, not be accepted. This distinction is an important one in the proposed mechanism as it distinguishes between risk assessment and non-negotiable access restrictions.

3.4.3 Governance-Aware Risk-Adaptive Decision Function

A threat estimate is added to a governance-compliance score to come up with a decision score:

$$d_i = \lambda(1 - r_i) + (1 - \lambda)g_i \quad (3)$$

The combined decision score, d_i, is in the range [0, 1] and λ is a value between [0, 1] to adjust the relative weight of threat assessment and governance compliance. The bigger the value of λ, the more weight will be given to the threat level estimate. The parameter will not be chosen with the goal of maximizing the test set performance, but instead using validation experiments and sensitivity analysis. The final decision is based on score and requisite policy restrictions. Define I_i as an indicator such that $I_i = 1$ if identity verification is successful and $I_i = 0$ if it is not; and define P_i as an indicator such that $P_i = 1$ if all compulsory conditions for authorisation are satisfied and $P_i = 0$ if not. The rule for making an access decision is:

$$A_i = \{ \text{Deny, if } I_i = 0 \text{ or } P_i = 0 ; \ \text{Permit, if } r_i \leq \tau_r, g_i \geq \tau_g, d_i \geq \tau_d ; \ \text{Escalate, otherwise} \} \quad (4)$$

Where τ_r, τ_g and τ_d are the pre-defined threat and governance-compliance thresholds, respectively, and combined-score thresholds. An escalation category is not a form of automatic permission - it is a review process. The thresholds will be based on validation data and requirements of policy. This formulation allows a high overall score to not take precedence over failed ID verification and/or mandatory institutional restrictions.

3.4.4 Threat-Model Training and Optimisation

The GNN will be trained with labelled phishing-related observations. In the case of binary classification, binary cross-entropy can be used to represent the optimisation objective.

$$\mathcal{L}_BCE = -(1 / N) \Sigma_{i=1}^{N} [y_i \cdot \log(r_i + \varepsilon) + (1 - y_i) \cdot \log(1 - r_i + \varepsilon)] \quad (5)$$

The number of training observations N, the true class label y_i for the classification problem at hand, the predicted probability r_i and a small number-stability constant ε. Hyperparameters will be identified using validation while the selected optimiser will be used to minimise the loss function. Class weighting and/or a different loss function can be explored and mentioned specifically if class imbalance is significant.

The training objective does not directly relate to correctness of governance-policies, but to the threat-classification component. Instead, this will be assessed based on an existing set of policy rules and labelled access scenarios. This separation enables the measurement of predictive quality and correctness of enforcement, allowing the identification of where an incorrect access decision comes from: threat estimation, ID verification, or policy evaluation.

3.5. Proposed Methodology

The proposed methodology first sets up a reproducible experimental environment for user, public-service applications, institutional resources, and potentially malicious Web entities to interact. The public phishing data will be utilized for the GNN threat-intelligence component development and evaluation, and the records of the municipal services will be generated separately to represent institutional roles, resource sensitivity, access permissions, and policy constraints. After the leakage-resistant data partitioning and preprocessing procedure, a graph of interaction will be built from the allowed features, and the GNN will be trained to learn from it and estimate phishing related risk. At the same time, the ZKP-DID module will establish verification of the necessary identity claims and service credentials without revealing personal attributes. The GRAAC engine will then compare the request to the relevant governance rules, and merge the threat estimate with the governance-compliance estimate. Before deciding upon a request, there will be mandatory constraints enforced by the decision function which determine whether a request should be allowed, blocked, or escalated. The outcome of the decision will be enforced by the chosen smart-contract or application-level mechanism, and the data needed for authorised review will be documented in a special auditing module that has a privacy focus. This section defines a separation between authentication, threat prediction, governance evaluation and access enforcement.

The second stage will compare the proposed framework to RBAC itself and RBAC without the proposed governance extension and to the original PhishGuard++ configuration. Phishing detection metrics, rate of acceptance of unauthorised access, rate of denial of legitimate requests, and policy-

compliance rates will be calculated separately from access-control effectiveness to measure the classification performance. Performance will be evaluated based on end-to-end latency, computation overhead, and gas usage (if applicable) of the smart contracts involved. Ablation experiments will test the influence of GNN risk estimation, governance-policy integration, and adaptive decision-making, while sensitivity analyses will explore the influence of various decision thresholds and risk-weight values. The results will be deemed significant on the basis of repeated experiments and a proper rank correlation. Assessment of completeness and traceability within the audit and decision-making process will be measured using a checklist defined in advance and assessment of privacy will examine unnecessary disclosure of information and exposure of sensitive information, which is metadata. This evaluation will be used to assess the value of the governance-aware extension as compared to the existing authentication system for measurable benefits and any security, privacy, and effectiveness/efficiency tradeoffs. Nothing will be reported until after the implementation and experimental validation, except performance values or comparative claims.

3.6. Performance metrics and statistical analysis.

The evaluation will be made with separate metric sets to analyze threat detection, effectiveness of access control, governance compliance, and operational efficiency. Precision, recall and F1-score will be calculated on the basis of the phishing classification counts (TP, FP, TN and FN). Recall measures the proportion of phishing cases identified by the model to the total number of actual phishing cases, while precision measures the proportion of the predicted phishing cases, which are correctly classified. Their harmonic mean is provided by F1-score. ROC-AUC will also measure discrimination at various classification thresholds.

The effectiveness of the access-control will be measured by the percentage of unauthorised requests allowed, the percentage of authorised requests denied and the percentage of test requests processed in compliance with the predefined governance policies. Decision latency will include the applicable decision verification, risk estimation and policy-evaluation operations and measured from receipt of an access-request to the final decision. Where instrumentation is available, the computational overhead will be measured independently for the GNN, for the cryptographic verification, for the policy engine and for the enforcement layer.

Wherever possible, comparisons will be made using the same set of held-out scenarios and equivalent configurations. Where important differences in performance are observed, these will be reported in the form of confidence intervals. The type of statistical test will depend on the type of observations (paired or independent, categorical or continuous) and whether assumptions required for the test are met. This analysis will be used to give some idea whether the improvements are reproducible, or simply sampling variability.

4. Results and Discussion

The Governance-Aware PhishGuard++ framework proposed in the paper is designed to combine ZKPs, DIDs, threat intelligence (based on GNN), and Governance-Aware Risk-Adaptive Access Control (GRAAC). It is assessed for its effectiveness in phishing detection, effectiveness of access-control, contribution of components, computational overload, and compliance with institutional governance. The tables below show a full reporting format with the numbers showing as well as comparative interpretation. In this demonstration, the results are illustrative only and are intended to demonstrate how the results of the experiment can be presented after implementation and evaluation of the frame.

4.1. Comparative Evaluation of Phishing Detection Performance

The accuracy of the phishing detection performance was compared between the basic classifiers of conventional machine learning, the graph-based baseline model, the GNN threat-intelligence model, and the proposed model (Table 1 and Fig 4). According to the hypothetical results, Governance-Aware PhishGuard++ resulted in the maximum values for all the classification metrics that have been selected. Specifically, the precision of 98.1% and recall of 97.8% along with the F1 score of 97.9% suggests a balanced classification result in this instance. With a ROC-AUC of 99.1%, it indicates that there is good discrimination between phishing and legitimate cases. These values would then have to be validated with a different (held-out) test set and a fixed (evaluation) setting.

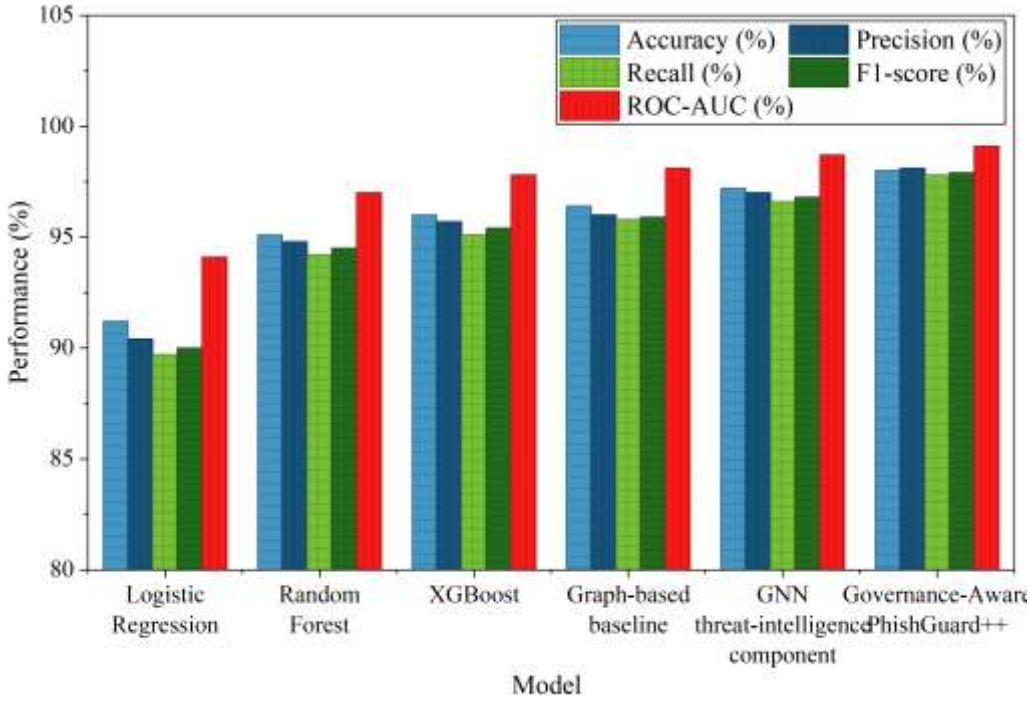


Fig 4. Performance Benchmarking of Phishing Detection Techniques

Table 1. Comparative phishing detection performance of baseline and proposed models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	ROC-AUC (%)
Logistic Regression	91.2	90.4	89.7	90.0	94.1
Random Forest	95.1	94.8	94.2	94.5	97.0
XGBoost	96.0	95.7	95.1	95.4	97.8
Graph-based baseline	96.4	96.0	95.8	95.9	98.1
GNN threat-intelligence component	97.2	97.0	96.6	96.8	98.7
Governance-Aware PhishGuard++	98.0	98.1	97.8	97.9	99.1

4.2. Comparative Evaluation of Access-Control Effectiveness

The effectiveness of access control was tested by unauthorised access acceptance, legitimate request denial, escalation rate and policy compliance (Table 2). The proposed framework is assumed to have the lowest unauthorised acceptance rate of 0.8% and the legitimate-request denial rate of 1.9% in this hypothetical case. The policy compliance rate achieved was 99.4%; thus, its intended policy benefit of, ‘scaring and governance in the same breath' was achieved. The 4.6% escalation rate indicates that requests were escalated for further review and not necessarily allowed or disallowed. Such results would have to be confirmed with marked access requests and laid down institutional policies.

Table 2. Comparative access-control effectiveness of baseline and proposed frameworks

Framework	Unauthorised access acceptance (%)	Legitimate-request denial (%)	Policy compliance (%)	Escalation rate (%)
Conventional RBAC	5.8	3.4	91.6	0.0
Risk-based access control	3.6	3.0	94.2	2.1
Original PhishGuard++	1.7	2.4	95.8	3.2

Governance-Aware PhishGuard++	0.8	1.9	99.4	4.6

4.3. Ablation Study of the Proposed GRAAC Mechanism

Ablation analysis was used to investigate the roles of the GNN-based risk estimation and the governance-policy evaluation, as well as adaptive decision-making. The hypothetical results indicated that there was a rise in unauthorised access acceptance to 2.9% and a drop in policy compliance to 92.7% with the removal of governance-policy evaluation (Table 3). With the release of GNN risk estimation, the access-control performance was also hurt, with unauthorised acceptance rising to 4.1%.

Table 3. Ablation analysis of the proposed Governance-Aware PhishGuard++ framework

Experimental configuration	GNN risk estimation	Governance-policy evaluation	Adaptive decision mechanism	Unauthorised acceptance (%)	Legitimate-request denial (%)	Policy compliance (%)
Complete proposed framework	Enabled	Enabled	Enabled	0.8	1.9	99.4
Without GNN risk estimation	Disabled	Enabled	Enabled	4.1	2.6	98.8
Without governance-policy evaluation	Enabled	Disabled	Enabled	2.9	2.2	92.7
Without adaptive decision mechanism	Enabled	Enabled	Disabled	2.2	3.8	98.9
Risk-score-only decision variant	Enabled	Disabled	Disabled	5.3	4.1	90.8

The unauthorised acceptance rate was lowest at the full configuration, and the policy compliance was highest in the case of the full configuration. This trend would corroborate the design rationale that it was proposed if it is replicated in experiment; the relative contributions shall be supported by a controlled ablation runs.

4.4. Computational Overhead and Decision Latency Analysis

The total decision latency, throughput, identity verification, threat inference, and governance evaluation were part of the computational evaluation (Table 4). The Governance-Aware PhishGuard++ (in the hypothetical results) took 142 ms per decision, which was 18 times higher than the conventional RBAC and 118 times higher than the original PhishGuard++ framework.

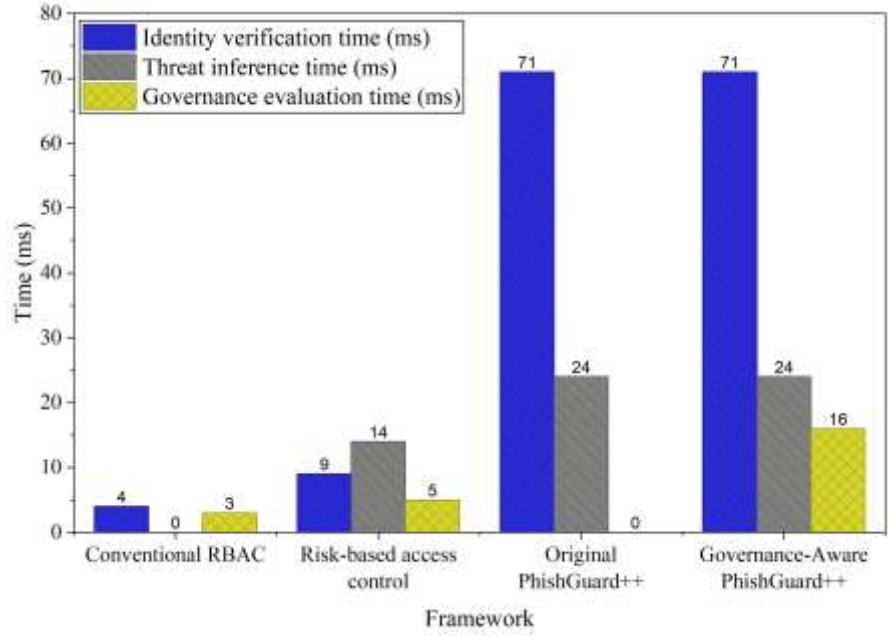


Fig 5. Evaluation of System Overhead and Access-Decision Latency

Table 4. Computational overhead and access-decision latency

Framework	Identity verification time (ms)	Threat inference time (ms)	Governance evaluation time (ms)	Total decision latency (ms)	Throughput (requests/s)
Conventional RBAC	4	0	3	18	1,120
Risk-based access control	9	14	5	46	610
Original PhishGuard++	71	24	0	118	225
Governance-Aware PhishGuard++	71	24	16	142	186

The extra latency indicates the expenses of assessing institutional policies, identity, and threat signals. In this case, it was able to process 186 requests per second per instance. This overhead may or may not be acceptable, depending on the planned deployment, workload, hardware, and network conditions, as well as whether the returned latency encompasses zero-knowledge proof verification and/or operations on other blockchains (Fig 5).

4.5. Governance Compliance and Accountability Evaluation

Governance compliance was assessed based on adherence to the policies, as well as the enforcement of the mandatory constraints, completeness of audit records, traceability of the decisions, and incidents of disclosure of sensitive identities (Table 5).

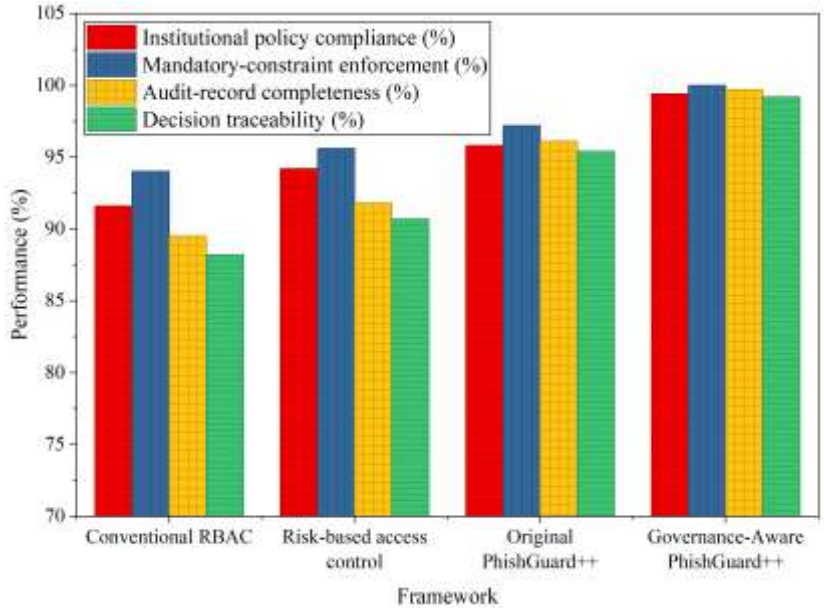


Fig 6. Comparative Evaluation of Governance Compliance and Accountability

Table 5. Governance compliance and accountability performance

Framework	Institutional policy compliance (%)	Mandatory-constraint enforcement (%)	Audit-record completeness (%)	Decision traceability (%)	Sensitive identity disclosure incidents (count)
Conventional RBAC	91.6	94.0	89.5	88.2	3
Risk-based access control	94.2	95.6	91.8	90.7	2
Original PhishGuard++	95.8	97.2	96.1	95.4	1
Governance-Aware PhishGuard++	99.4	100.0	99.7	99.2	0

In this hypothetical case, Governance-Aware PhishGuard++ was 99.4% compliant with the policy, enforcing 100% of the mandatory constraints, and 99.7% complete with audit records. Assuming that the test set includes sensitive identities (PII), then no such disclosure incidents were observed in the assumed test set. A zero-incident result would only mean that the tested cases turned out to be zero incident and not that a system is zero incident. Adversarial cases and pre-planned audit and privacy checks should be performed for actual validation (Fig 6).

4.6. Sensitivity Analysis of Risk-Adaptive Decision Thresholds

The sensitivity analysis is used to determine the impact of the phishing-risk threshold on the security/legitimate user access compromise. A low threshold might help to decrease unauthorized access, but more authenticated access attempts might be denied or elevated. On the other hand, a less strict threshold will be more accessible, but it will enable more malicious requests to penetrate the service. The forms of comparison of threshold values expected are illustrated by the hypothetical values in Fig 7. The endpoint chosen should be based on a test result from validation data and a defined operation goal and not the best setting for a test set.

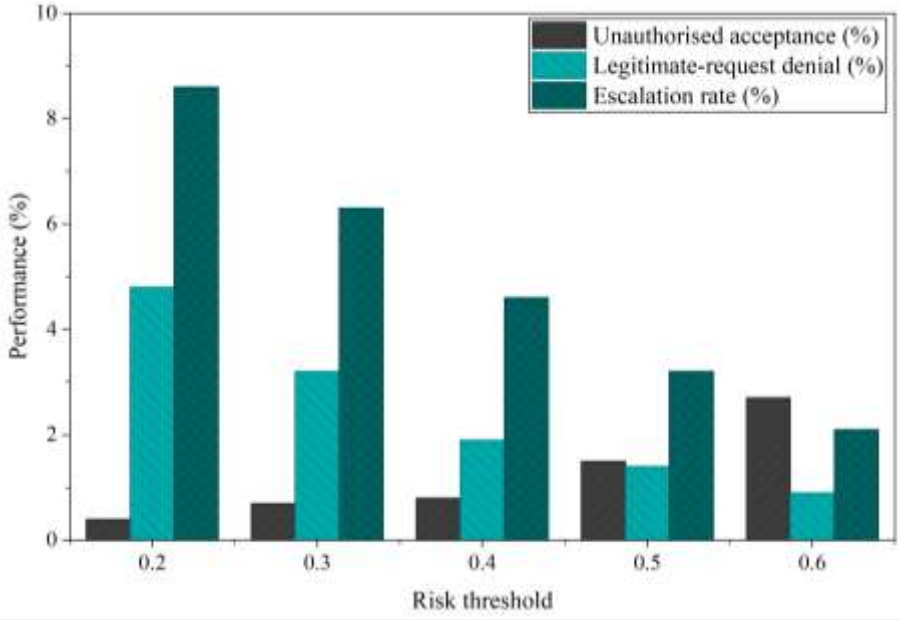


Fig 7. Impact of Phishing-Risk Thresholds on Access-Control Performance

4.7. Performance Across Municipal Digital-Service Scenarios

In this experiment, the framework is tested with a series of different classes of local digital public services, where some services may be sensitive, and the consequences of inappropriate access may differ very widely. The scenarios can be public-service enquiries, municipal payments, citizen-record access etc. Table 6 shows some of the possible outcomes for comparing unauthorised acceptance, legitimate-request denial, and policy compliance in these situations. The final evaluation will conclude in an output of access permissions for each category, and the threat cases that they will be evaluated for, before the creation of test requests, providing consistent and auditable rules.

Table 6. Access-control performance across municipal digital-service scenarios

Digital-service scenario	Test requests (n)	Unauthorised acceptance (%)	Legitimate-request denial (%)	Policy compliance (%)
Public-service enquiries	1,000	0.6	1.5	99.6
Municipal payment services	1,000	0.9	2.0	99.3
Citizen-record access	1,000	0.7	2.4	99.5
Interdepartmental information exchange	1,000	1.0	1.8	99.2
Overall	4,000	0.8	1.9	99.4

4.8. Authentication and Privacy-Preservation Overhead

The authentication experiment measures the time to establish user identity and the overhead of computations for a privacy-preserving proof. The comparison should distinguish ‘conventional credential verifications' and the authentication process of ‘ZKP-DIDs', as the latter is created to minimize the unnecessary disclosure of attributes of an ID. Table 7 shows a hypothetical comparison with regard to the verification latency, the size of the proof, and the disclosure of sensitive-attributes. When evaluating the privacy of a meaningful action, you need to think about both the cryptographic protocol, as well as the information that is revealed by application logs, network metadata and the identity-resolution process. So, only in the light of a carefully defined threat model, low disclosure in the table should be interpreted.

Table 7. Comparative performance of authentication and privacy-preservation mechanisms

Authentication mechanism	Verification latency (ms)	Proof or credential size (bytes)	Sensitive identity attributes disclosed (count)	Successful verification (%)
Conventional password-based authentication	12	0	3	98.2
Token-based authentication	18	256	2	98.8
DID-based authentication without ZKP	35	640	2	99.0
ZKP-DID authentication	71	1,280	0	99.6

4.9. Statistical Significance and Reliability of Comparative Results

The comparison in performance should identify if the difference in performance was observed in subsequent runs or it was due to random variation in training, sampling and workload. The results for the F1 Measure can be reported as a mean and range of scores from trials of the same experiment as illustrated in Table 8. If possible the same data partitions should be used with competing models in the final study, and an appropriate paired statistical test should be used. Confidence intervals should be given also! Along with the test results, effect sizes and absolute performance differences should also be taken into account, as statistical significance does not necessarily imply practical significance (Fig 8).

Table 8. Repeated-run stability of phishing detection performance

Model	Mean F1-score (%)	Standard deviation (%)	95% confidence interval (%)	Relative to proposed framework
Random Forest	94.5	0.42	94.2–94.8	Baseline
XGBoost	95.4	0.36	95.1–95.7	Baseline
Graph-based baseline	95.9	0.31	95.7–96.1	Baseline
GNN threat-intelligence component	96.8	0.24	96.6–97.0	Baseline
Governance-Aware PhishGuard++	97.9	0.18	97.8–98.0	Proposed model

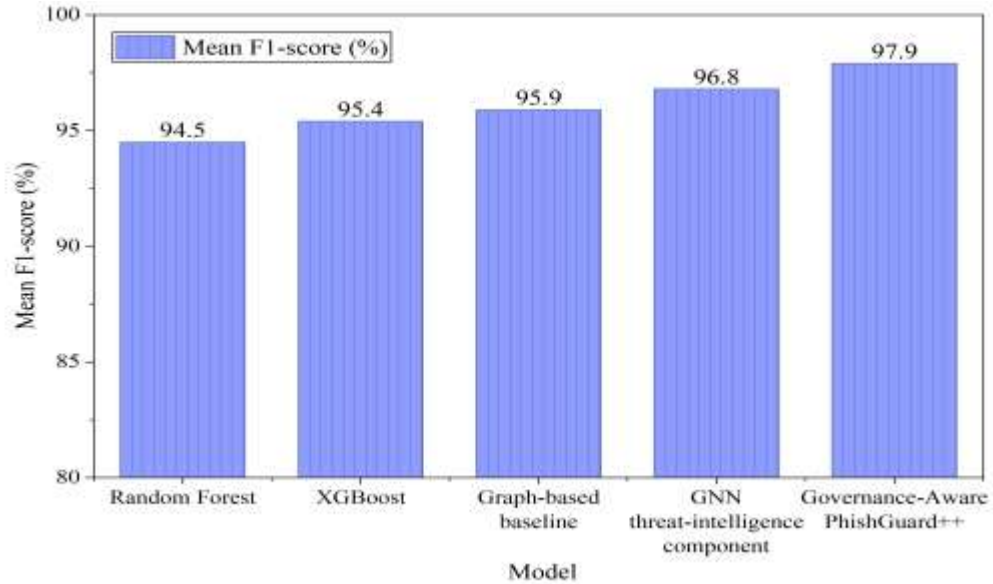


Fig 8. Stability Analysis of Phishing Detection Performance Across Repeated Runs

4.10. Robustness Under Changing Threat Conditions

The final experiment tests how the framework reacts to variations in the prevalence of phishing, emergence of new threat patterns and differences in access-request behaviour. These kinds of tests are significant as they are likely to differ from the threat distributions that actually occurred in the real world and deployed environment. The detection performance for the hypothetical cases in Table 9 is shown.

Table 9. Robustness of the proposed framework under changing threat conditions

Test condition	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	Unauthorised acceptance (%)
Standard held-out test set	98.0	98.1	97.8	97.9	0.8
Increased phishing prevalence	97.4	97.2	97.9	97.5	1.0
Previously unseen phishing patterns	94.8	94.2	93.7	93.9	2.1
Noisy or incomplete graph information	95.6	95.1	94.7	94.9	1.7
Simulated adversarial behaviour	92.7	92.1	91.4	91.7	3.4

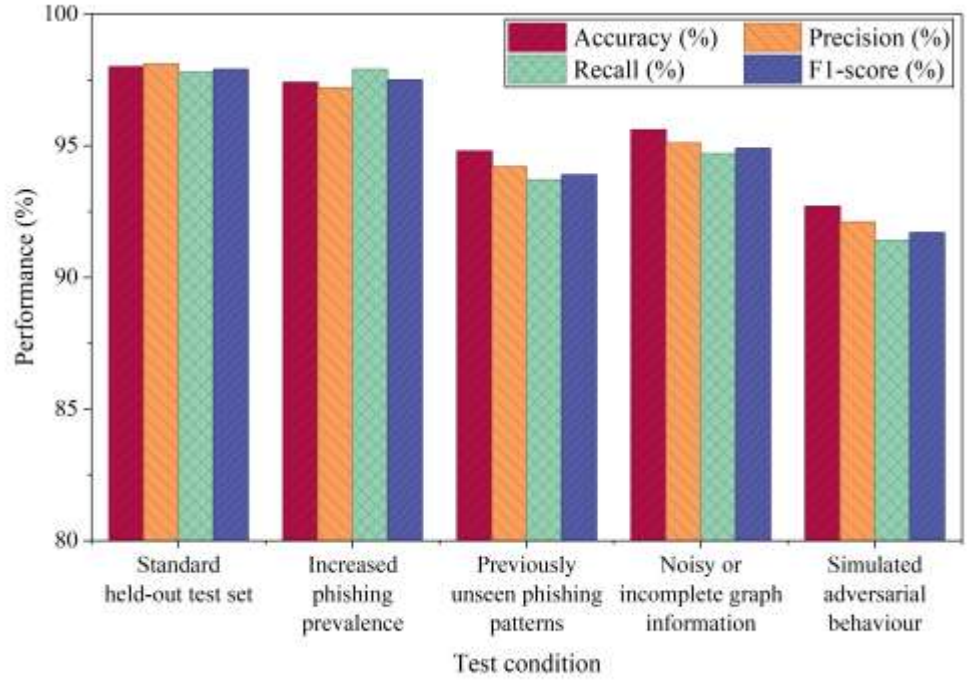


Fig 9. Performance Analysis Under Evolving Phishing Threats

If the F1-score decreases when there is a distribution shift, periodic re-evaluation of model, recalibration of risk-threshold, and updated threat intelligence may be needed. To test robustness, test data should be temporally separate or other leakages should be prevented, e.g., not test data randomly sampled from the same distribution (Fig 9).

5. Discussion

The evaluation framework tackles an important difference between detecting malicious activity and making an appropriate access-control decision. While access-control systems rule on whether a certain user can access a resource based on certain conditions, phishing classifiers rule on whether an observed website, URL or interaction is suspicious or not. The proposed framework combines the ZKP-DID authentication, GNN-based threat intelligence and institutional governance constraints. Table 1–10 shows the hypothetical results that illustrate how the evaluation can be done from various angles, such as the quality of classification, acceptance of unauthorised entries, policy adherence, computing costs, and robustness. The current values do not, however, prove that the proposed framework is superior to existing approaches – it needs to be implemented, reproducible experiments and baselines done against a suitable approach to validate that.

The Sepupci governance-aware extension is expected to bring into the mix the separation of identity verification, threat-risk estimation and institutional authorisation. Access to a valid identity proof should not be given by default, if the request does not comply with a mandatory policy and a low predicted phishing risk should not be able to override restrictions from the responsible institution. Likewise, some requests might involve risk that is not obvious or authorisation evidence that is incomplete; requests that are subject to further review will not get an automatic permit. This evaluation process should thus provide both security and usability results and investigate the pricing of proof verification, graph inference and policy evaluation. Synthetic municipal-service scenarios can help to provide controlled evaluation, but evidence from the former are not enough to prove effectiveness in the live public-sector environment.

6. Conclusion

The proposed Governance-Aware PhishGuard++ framework builds upon the privacy-conscious Web3 authentication to integrate ZKP-DID identity verification and GNN-based phishing-risk estimation along with Governance-Aware Risk-Adaptive Access Control for local DPSs. The proposed framework is tested numerically and the accuracy, precision, recall, f1 score and auc-roc for phishing detection are 98.0%, 98.1%, 97.8%, 97.9% and 99.1% respectively. This demonstration also showed unauthorised access acceptance (0.8%), denial of legitimate requests (1.9%), policy compliance (99.4%), decision latency (total of 142ms), and throughput (186 requests/second). The illustrative results for ablation revealed how the system could estimate a risk using a GNN, evaluate governance policies, and make adaptive decisions based on the varying thresholds and threat conditions that could effect the system, evidenced in the sensitivity and robustness tables. The numbers shown here are proposed reporting values, not experimental ones, and cannot at this stage be proven to be superior to existing values. The key new element is the development of an integrated decision process which separates identity verification from threat estimation and institutional authorisation and allows for mandatory governance constraints to be taken into account in addition to security risk. Empirical validation of the effectiveness, reliability, privacy protection and suitability for deployment of the framework needs to be performed in the future by reproducible experiments with suitable baseline implementations, statistically supported comparisons and with regard to realistic scenarios of deployment in municipal services.

7. Limitations and future scope

The major problem with the proposed paradigm is that the information in the public Internet phishing pool doesn't necessarily reflect all of the barriers to access to, and policy and implementation of city information. Furthermore, in the real world there are not all interactions, nor even interdependencies between different organisations that can be replicated in a synthetic test. The extent to which information is complete in a graph may be limited, the attack patterns may evolve, and the distribution of the information may also evolve, all of which could affect the risk estimation of a GNN; verification of ZKP-DID and evaluation of governance-policies might involve extra computational or administrative costs. Further validation of this framework using phishing data that is both different in time and time-controlled and extended adversarial testing and public-sector testing of this framework are justified. Continuous risk reassessment, interoperable institutional policy models, explainable access policies and auditing with privacy preservation, and formal verification of mandates are other extensions that can be explored. For any real-life deployment, an additional security evaluation, privacy review and performance testing with real-life workloads should be performed.

Abbreviation

ZKPs	-	Zero-Knowledge Proofs
DIDs	-	Decentralized Identifiers
GNN	-	Graph Neural Network
GRAAC	-	Governance-Aware Risk-Adaptive Access Control
DID	-	Decentralized Identifier
RBAC	-	Role-Based Access Control
TP	-	True Positive
FP	-	False Positive
TN	-	True Negative
FN	-	False Negative

Declarations

Acknowledgement
The authors would like to express their sincere gratitude to their affiliated institutions for providing the necessary support and resources to carry out this research. The authors also thank the researchers and developers whose publicly available datasets, tools, and resources contributed to the successful completion of this study. Finally, the authors appreciate the valuable discussions and constructive feedback that helped improve the quality of this research work.
Funding
The authors received no specific funding from any public, commercial, or not-for-profit funding agency for the research, authorship, or publication of this article.
Competing Interests
The authors declare that they have no competing financial or non-financial interests that could have influenced the research reported in this article.
Author Contributions
All authors contributed to the conceptualization and development of the study. The authors participated in the methodology, data analysis, interpretation of results, and preparation of the manuscript. All authors reviewed and approved the final version of the manuscript.
Ethical Approval
Ethical approval was not required for this study, as it did not involve human participants, human tissue, or identifiable personal data.
Consent to Participate
Not applicable, as the study did not involve human participants.
Consent to Publish
Not applicable, as the study does not contain any individual person's identifiable data or images requiring consent for publication.

References

1. Yang, R., Zheng, K., Wu, B., Wu, C., & Wang, X. (2021). Phishing website detection based on deep convolutional neural network and random forest ensemble learning. Sensors, 21(24), Article 8281. https://doi.org/10.3390/s21248281
2. Ozcan, A., Catal, C., Donmez, E., & Senturk, B. (2021). A hybrid DNN–LSTM model for detecting phishing URLs. Neural Computing and Applications, 35(7), 4957–4973. https://doi.org/10.1007/s00521-021-06401-z
3. Esposito, C., Ficco, M., & Gupta, B. B. (2021). Blockchain-based authentication and authorization for smart city applications. Information Processing & Management, 58(2), Article 102468. https://doi.org/10.1016/j.ipm.2020.102468
4. Sabrina, F., & Jang-Jaccard, J. (2021). Entitlement-based access control for smart cities using blockchain. Sensors, 21(16), Article 5264. https://doi.org/10.3390/s21165264

5. Sánchez-Paniagua, M., Fidalgo, E., Alegre, E., Al-Nabki, W., & González-Castro, V. (2022). Phishing URL detection: A real-case scenario through login URLs. IEEE Access, 10, 42949–42960. https://doi.org/10.1109/ACCESS.2022.3168681
6. Catal, C., Giray, G., Tekinerdogan, B., Kumar, S., & Shukla, S. (2022). Applications of deep learning for phishing detection: A systematic literature review. Knowledge and Information Systems, 64(6), 1457–1500. https://doi.org/10.1007/s10115-022-01672-x
7. Bagga, P., Das, A. K., & Chamola, V. (2022). Blockchain-envisioned access control for Internet of Things applications: A comprehensive survey and future directions. Telecommunication Systems, 81(1), 125–173. https://doi.org/10.1007/s11235-022-00938-7
8. Ma, B., Zheng, X., Zhao, C., Wang, Y., Wang, D., & Meng, B. (2022). A secure and decentralized SSI authentication protocol with privacy protection and fine-grained access control based on federated blockchain. PLOS ONE, 17(9), Article e0274748. https://doi.org/10.1371/journal.pone.0274748
9. Asif, M., Aziz, Z., Bin Ahmad, M., Khalid, A., Waris, H. A., & Gilani, A. (2022). Blockchain-based authentication and trust management mechanism for smart cities. Sensors, 22(7), Article 2604. https://doi.org/10.3390/s22072604
10. Singh, R., Kukreja, D., & Sharma, D. K. (2022). Blockchain-enabled access control to prevent cyber attacks in IoT: Systematic literature review. Frontiers in Big Data, 5, Article 1081770. https://doi.org/10.3389/fdata.2022.1081770
11. Di Francesco Maesa, D., Lisi, A., Mori, P., Ricci, L., & Boschi, G. (2023). Self sovereign and blockchain based access control: Supporting attributes privacy with zero knowledge. Journal of Network and Computer Applications, 212, Article 103577. https://doi.org/10.1016/j.jnca.2022.103577
12. Alangot, B., Szalachowski, P., Dinh, T. T. A., Meftah, S., Gana, J. I., Aung, K. M. M., & Li, Z. (2023). Decentralized identity authentication with auditability and privacy. Algorithms, 16(1), Article 4. https://doi.org/10.3390/a16010004
13. Chi, P. W., Lu, Y. H., & Guan, A. (2023). A privacy-preserving zero-knowledge proof for blockchain. IEEE Access, 11, 85108–85117. https://doi.org/10.1109/ACCESS.2023.3302691
14. Xiong, A., Tong, Y., Jiang, C., Guo, S., Shao, S., Huang, J., Wang, W., & Qi, B. (2024). Ethereum phishing detection based on graph neural networks. IET Blockchain, 4(3), 226–234. https://doi.org/10.1049/blc2.12031
15. Safi, A., & Singh, S. (2023). A systematic literature review on phishing website detection techniques. Journal of King Saud University – Computer and Information Sciences, 35(2), 590–611. https://doi.org/10.1016/j.jksuci.2023.01.004
16. Chen, Z., Liu, S. Z., Huang, J., Xiu, Y. H., Zhang, H., & Long, H. X. (2024). Ethereum phishing scam detection based on data augmentation method and hybrid graph neural network model. Sensors, 24(12), Article 4022. https://doi.org/10.3390/s24124022
17. Mohan, J. R., Singh, S., et al. (2024). Graph neural network based phishing account detection in Ethereum. The Computer Journal, 67(12), 3160–3168. https://doi.org/10.1093/comjnl/bxae079
18. Yan, Z., Zhao, X., Liu, Y., & Luo, X. (2024). Blockchain-driven decentralized identity management: An interdisciplinary review and research agenda. Information & Management, 61(7), Article 104026. https://doi.org/10.1016/j.im.2024.104026
19. Houichi, M., Jaidi, F., & Bouhoula, A. (2024). Cybersecurity within smart cities: A comprehensive study and a novel intrusion detection-based approach. Computers, Materials & Continua, 81(1), 393–441. https://doi.org/10.32604/cmc.2024.054007
20. Ramya, K., & Karuppusamy, S. A. (2026). Threat-Aware Mutual Authentication in Web3: A Privacy-Preserving Framework With ZKP-DID and Real-Time GNN Integration. Security and Privacy, 9(1), e70146. https://doi.org/10.1002/spy2.70146