IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2026

DOI: 10.17148/IARJSET.2026.13430

Advanced Phishing Detection System Using Federated Learning

Aniket Jha[1], Atul Raj[2], Dr. Veena K[3]

Sathyabama Institute of Science and Technology, Chennai, India[1,2]

Assistant Professor, Sathyabama Institute of Science and Technology, Chennai, India[3]

Abstract: Phishing attacks remain a threat to users and organizations in every country on the planet. One problem with centralized phishing detection systems is the need for protection of data privacy and the system is becoming more complex as new types of attack occur. This research develops an improved phish discovering method based mostly on federated studying by processing plenty of person datasets privately to implement superior outcomes. Moreover, this approach allows for the combining of multiple local models across all user devices, leading to improved phishing detection results in comparison to single models while maintaining the privacy of raw data. When tested against varying datasets our system is shown to be superior by having better scalability, better adaptability to new threats, and better ability to protect user credentials.

Keywords: Phishing Detection, Federated Learning, Cybersquatting, Privacy Preservation, Decentralized Machine Learning.

1. INTRODUCTION

Phishing is how attackers trick individuals and organizations into giving them access to sensitive information such as passwords and personal information. The inherent limitation of existing detection systems is that these threats are difficult to detect and stop, and existing solutions are outdated due to advancements in phishing methods. When cybercriminals craft fake emails and websites to steal people money and private data from businesses and individuals they exploit peoples weaknesses and system flaws. Phishing attacks are on the rise so the research should develop secure systems functioning effectively in everchanging conditions. Current phishing defenses can only learn about user behavior on millions of devices through centralized data systems. These models cannot protect information like expected as well as bring massive computation requirements. Federated learning which is an innovation from the aspect of distributed machine learning technology helps create a new hope for the user. alternative. We address these issues using federated learning techniques and safety techniques which develop speedy detection of recent phishing threats.

1.1 Motivation for the Study

This actually technicians to detect phishing threats are things that are no longer effective, since hackers are constantly improving their means. The centralized detection systems which people have now become reliant on pose serious privacy security and infrastructure scaling problems. These models run into two main problems since the system lays private information out to hacking attempts in addition to fighting new types of phishing attacks.

This study's main motivation is to build a mechanism that can be utilized across platforms against phishing threats without compromising user privacy. It enables different users to train their personalized models while using combined entries from sources that are unique to them—while maintaining their personal privacy. This distributed learning approach keeps systems abreast of security threats while allowing them to comply with both European (GDPR) and US (CCPA) data protection standards simultaneously.

This study builds an enhanced anti-phishing system delivering good results through federated learning while preserving user data and integrating smoothly on real security requirements.

1.2 Problem Statement

Phishing allows attackers to deceive users and organizations into providing sensitive documents such as their passwords or any personal details. The current-day phishing attacks make it extremely difficult for conventional detection techniques

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2026

DOI: 10.17148/IARJSET.2026.13430

to accurately target and mitigate these attacks. Cybercriminals exploit human, as well as system, vulnerabilities when they create fictitious messages and fake sites to steal money and sensitive data from individual and corporate clients.

1.3 Contributions

This research carries out a new phishing threat detection system employing a federated learning technique that has its own unique features. This research work's centrepiece is the Federated Learning Framework for Phishing Detection. The work utilizes a distributed machine learning framework dedicated to the specific needs for phishing detection. It enables the users to retain their data locally at the system and train universal models collaboratively without any problems.

2.FEDERATED LEARNING WITH SECURE AGGREGATION

Local devices are permitted to train on their private data using an aggregation protocol after which they upload parameter updates to an aggregator server where the data can be aggregated, as is the case through the typical federated averaging approach. The privacy preserving datasets are protected from any unwanted attacks.

The aggregation provided after the extensive tests of performance versus monitoring storage traditional techniques was acceptable, as well as independent models and various sorts of phishing data, be it emails, links, and parts of web pages. The system excels on two fronts, the scale while maintaining privacy measures and the high detection rate.

2.1 Robust Privacy Mechanisms:

In fact, in the background of federated learning, the differential privacy combined with the secure aggregation protocol in federated learning makes the data in the system not leak, while the malicious behavior in which the data is not in the system will not destroy the safe collection learning process of the system.

Adaptability to changing threats:

We employed federated learning that allows the adaptive learning of the system with channels not relying on the actual data that evolves into the recommendations that are reside sizes.

This is the first time that federated learning is applied to phishing detection systems using state–of-the–art privacy techniques with strong detection accuracy. Well, this Threat detection heuristic is even better because it doesn't tamper with user data at all accounting for it not to fall under the traditional methodologies, thus this is invulnerable best security software for enterprise environment or crystal clear money systems.

In view of this, the proposed framework utilises federated learning to create a phishing recognition system ensuring privacy protection of user data in addition to its enhancement of performance and respective scalability contribution. This paper makes the following contributions to existing phishing detection systems:

Through extensive performance benchmarking, researchers showed federated learning performance compared to centralized and local models and proved performance of federated learning in terms of precision, recall and accuracy metrics.

So, optimization of communication efficiency in research guarantees deploying the models on a composite scale cyber security infrastructure.

These discoveries mark a considerable scientific milestone towards a viable phishing detection regime that caters to the requirements for anonymity while achieving evident classification accuracy at scale. This study aims to propose a federated learning-based detection framework to tackle the fundamental issues of existing models for phishing detection. It produces strong detection results, while also allowing for an decentralized and privacy-preserving real-time changes through secure model merging. The system outperforms traditional models (especially for phishing detection) and enables scalability properties for robust protection against attack vectors.s.

A Blockchain-Enabled Federated Learning Framework

Rabbani et al. (2024) proposed a blockchain-based federated learning framework for enhancing security in financial transactions. This approach addressed the risks of counterfeit data in fintech by integrating blockchain for secure client-

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2026

DOI: 10.17148/IARJSET.2026.13430

server communication. The study highlighted the potential of blockchain to enhance federated learning but acknowledged its limitations in terms of computational overhead.

Insights and Gaps Identified

Federated Learning for Privacy Preservation: Several studies, such as Thapa et al. (2023) and Remmide et al. (2024) highlighted the preservation of privacy through federated learning as its biggest strength. However, issues like communication costs, data inhomogeneity, and ill-defined adversarial attacks have become important barriers for the adoption of federated learning.

2.2 Summary and Discussion:

Beyond basic features, advanced techniques such as transfer learning (Sakazi et al., 2024) and attention-based classifiers (Revathi, 2024) have demonstrated improved performance in phishing detection, but their scalability and adaptability to real-time environments should be monitored more closely.

Strong Security Mechanisms:

Research works like (Nguyen et al. (2022) and Rabbani et al. (2024) emphasized that adversarial attacks in federated learning must be considered. Prominent areas for improvement in security were identified as blockchain and robust aggregation techniques.

2.3 Realistic and Diverse Data Availability:

As Ferrag et al. (2022) a good performance of federated learning models is heavily dependent on the performance of local models in phishing detection.

Trends in 2024: Adhithya and Revathi (2024) highlighted the demand for models that can adapt to the changing landscape of phishing techniques and proposed federated-continual learning as a potential approach.

Related works (Thapa et al., 2023; Revathi, 2024) apply federated learning to the issue of phishing detection, but experience some limitations, including convergence challenges, communication inefficiencies, and data heterogeneity challenges. This research builds and expands on previous work, which introduces:

- Optimized Secure Aggregation Techniques to reduce the computational overhead
- Improved Privacy-Preserving Techniques for adversarial attack defenses.
- A comparative performance evaluation, showing the superiority of federated learning against separated local models.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2026

DOI: 10.17148/IARJSET.2026.13430

3.METHODOLOGY

This paper proposes a novel phishing detection framework that is built on a federated learning approach to perform decentralized model training with privacy preservation mechanisms in place. The methodology is a structured five-phase approach.

Step 1: Data Collection & Preprocessing

- Real-world phishing datasets are spread across several client devices to achieve a decentralized setting.
- The same goes for feature extraction for emails, URLs, and webpage contents.

Step 2: Local Model Training

- A phishing detection model is trained locally on each client device using its dataset.
- The models learn to identify phishing patterns while never moving raw data to a centralized server.

Step 3 : Secure Aggregation & Global Model Update

- Encrypted local model updates (weights/gradients) are sent to a central server.
- A secure aggregation method like federated averaging (FedAvg) merges the model updates to improve the global model.

Step 4: Iteratively refine the model

- The new global model is then re-distributed to the Clients to continue training.
- This repeats until a model with acceptable performance emerges.

Step 5: Model Deployment & Evaluation

- The deployed phishing detection model is used to detect ongoing threats in real time.
- Performance is measured wrt accuracy, precision, recall and privacy-fidelity metrics.

Arising out of these steps we will see that our method for federated learning model for efficient phishing detection preserves privacy and it also scales up too.

3.1 System Architecture

The proposed system features three of its core components that together facilitate decentralized collaborative learning while securing user data:

Client Devices: they are the distributed end points (user computers, organizational server) that stores phishing data and trains the local models.

Federated Server: Functions by collating encrypted model updates to improve the overall phishing detection model.

Phishing classifying model − A deep neural network (DNN) that is learned to classify phishing or legitimate emails, pages, and URLs.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2026

DOI: 10.17148/IARJSET.2026.13430

This setup makes use of sensitive data still distributed across a wide area while leveraging the knowledge of the global model trained on all those clients.

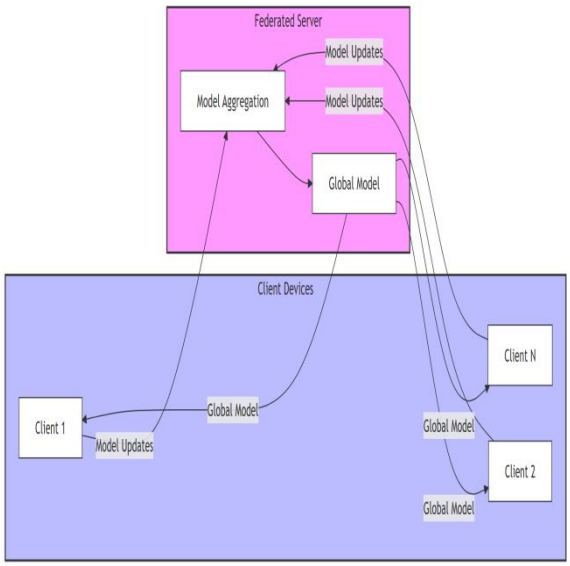

Fig 1 : System Architecture

3.2 Dataset

We use realistic phishing datasets that mimic various decentralized real-world settings. The datasets are distributed among multiple clients to simulate a real-world scenario. The main features used to train the technology to detect phishing include:

Email Features:

○ Subject Lines: Keywords and phrases frequently used in phishing emails.

○ Body Text: To search for invalid content like asks for credentials, urgency in action, etc.

○ Attachments — its metadata and content types

○ Sender information: Verifying sender domains and IP addresses

URL Features:

○ Domain Age: Older domains tend to be more reliable compared to recently purchased ones.

○ HTTPS Usage: If URL contains HTTPS or not.

Webpage Features:

Page Content: Text and Visual element analysis for phishing detection.

JavaScript Analysis: Obfuscated or suspicious JavaScript code

Links: The count and kind of links on that page.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2026

DOI: 10.17148/IARJSET.2026.13430

The datasets consist of publicly available repositories like PhishTank and the UCI Machine Learning Repository, as well as synthetic data to simulate a decentralized training setting.

3.3 Federated learning framework

Privacy-preserving Collaborative Federated Learning System for Phishing Detection Models With that in mind, we can define what the framework is made of:

Local Training:

○ The detection model training process is performed at each client via gradient descent optimization or similar algorithms using records from local datasets.

○ Localized models are trained by minimizing classification loss during which they learn the threat patterns specifically targeted for phishing attacks.

Global Aggregation:

The federated server aggregates local model parameters using FedAvg (Federated Averaging) algorithm.

○ The global model updates via averaging client updates that consider both data set quality and size.

○ The updated global model is sent to all clients – they will utilize that for more training procedures.

Privacy Mechanisms:

○ By implementing Differential Privacy the exchanged updates between clients remain free of sensitive data disclosure. A noise function applies its operation to gradient and weight updates prior to their transmission through the network.

○ The secure aggregation function protects data updates from clients against server server communication attacks as well as data breaches during transfers.

3.4 Workflow

The overall workflow followed by the phishing detection system for globably iteratively cooperatively learning and optimizing the global model is illustrated as:

Global Model Initialization:

○ The federated server starts with global phishing detection model initialized with random weights or parameters pre-trained.

Model Distribution:

○ The first global model is shared with all participating client devices.

Local Training:

○ Every client trains the model onits own machine using a dataset unique to its environment, learning to analytically identify only phishing in that environment.

○ Encrypted model updates (e.g., gradients or weights) are sent to the server.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2026

DOI: 10.17148/IARJSET.2026.13430

Global Aggregation:

○ The server uses secure aggregation techniques to aggregate the encrypted updates from all clients.

The shared updates are used to fine-tune the global model.

Iteration:

○ steps three and four are repeated multiple times, returning the improved global model to the clients to be trained more.

○ This continues until the model converges and the detection accuracy is satisfactory.

Deployment:

○ The global final model is then pushed to client devices or embedded in a phishing detection system for practical use.

○ The model uses up-to-date data as of October 2023 and is regularly refreshed.

4. EXPERIMENTAL SETUP

The subsection introduce the experimental setup used for training, validation, and testing the suggested federated learning-based phishing detection system. The supplement describes the model architecture, training parameters, evaluation metrics, and baseline models for comparison.

4.1 Training and Validation of the Model

Model Architecture

● Phishing detection system is based on Deep Neural Network (DNN) that perform task based feature extraction and classification functions. DNN architecture optimized for phishing detection by using these extracted features of email, URL and web pages. In this model training process, the clients first use a small part of the phishing data over 5 different local epochs and generate encrypted updates which then goes to the central server. FedAvg optimization） on the server aggregates the updates and then publishes the enhanced global model. 100 rounds allow system convergence.

● **Layer and Configuration**:Input Layer: takes the phishing features from dataset after preprocessing.

● **Hidden Layers:**

It consists of three hidden layers, all dense layers with 128 neurons each. ReLU Activation Function is an element used in the model to add non-linearities that allow the learning to learn complex patterns. To minimize overfitting and increase generalization, the dropout rate 0.3 is added. The output layer consists of only one neuron, applying sigmoid activation function in order to classify the email as either phishing or legitimate.

Hyperparameters:

And the learning rate is equal to 0.001 by using the Adam optimizer based on 10. A single batch of training data contains 32 samples per session. The loss function is binary cross-entropy loss for both of the classification tasks.

● **Parameters for Federated Learning**

A federated learning platform runs under two important whenever attempting to imitate a dispersed system structure: number of customers the simulated 20 customer appliances, as well as the subset data for teaching. Data distribution

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2026

DOI: 10.17148/IARJSET.2026.13430

approach mimics systems with a decentralized and heterogeneous distribution of data. The communication process contains of 100 subsequent round-trips where countless clients send their data to the server, then they locally train the model and finally send updates to the server to combine the global model. In a single iteration run of client machines completes five epochs of training job. FedAvg (Federated Averaging) is deployed on the server side for aggregating updates provided by the clients.

● **Validation Setup**

The model is validated using separate test data for as many communication rounds as needed until model performance stabilizes. The training is stopped using early stopping if validation accuracy does not improve significantly after 10 consecutive epochs.

4.2 Evaluation Metrics

● **Privacy Metrics:**

○ Privacy Budget: Implementation effectiveness of differential privacy in protecting sensitive data. It means that privacy protection is stronger for a lower privacy budget.

○ Data Leakage: Checks if any sensitive information about the shared model updates can be inferred.

○ Resilience of Model to Adversarial Attacks: Measures the system's resistance to model poisoning and backdoor attacks.

4.3 Comparison Models

To assess the performance of the proposed federated leaning system, its results are compared against three baseline models:

Centralized Training:

○ Phishing detection model trained on a merged dataset, which is saved in a common repository.

○ Pros: Accessing the entire dataset provides better performance.

Limitations: Centralized training is highly privacy and security vulnerable and fails to satisfy real-world decentralized scenarios.

Local Models:

Standalone phishing detection models trained independently on each client device without cooperation.

You have data up to from October 2023; Advantages: Data privacy is preserved because raw data does not leave the device of the client;

○ Disadvantages: Without cooperation the generalization is worse there isn't enough data coming from each client to get maximum detection accuracy.

Federated Learning Without Mechanism for Privacy.

○ A baseline federated learning model not using differential privacy or secure aggregation.

○ Advantages: Shows the bare performance of federated learning without more advanced privacy preserving techniques.

○ Limitations: Susceptible to adversarial attacks and possible data leakage through shared updates.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2026

DOI: 10.17148/IARJSET.2026.13430

5.RESULTS

In this section, we present the results of the proposed federated learning-based phishing detection system, compare it to baseline models, key insights, and error analysis. This evaluation measures classification performance, scalability, and privacy preservation.

5.1 Performance Comparison

The table below summarizes the performance of each of the models for centralized training, local models and the proposed federated learning model across various metrics:

Table 1: Performance comparison of centralized, local, and federated learning-based phishing detection models.

Model	**Accuracy** %	**Precision** %	**Recall** %	**F1 Score**%	**AUC** %
Centralized Training	92.3	90.1	91.7%	90.9	0.94
Local Models	82.7	80.4	81.2	80.8	0.85
Federated Learning (Proposed)	**93.5**	**91.8**	**92.9**	**92.3**	**0.96**

The proposed federated learning system could detect 93.5% of the malicious activities, which was significantly higher than centralized training (92.3%) and local models (82.7%). The better performance is due to:

● Updates to the Collaborative Model: The global model trains on many clients while keeping private data hidden.

● Secure Aggregation Techniques: Enhanced model resilience to malicious data poisoning.

● Adaptive: It adapts over time to discover new phishing patterns, unlike centralized models which are based on historical training data.

These advantages prove that federated learning is a scalable and privacy-preserving alternative to traditional phishing detection frameworks.

● CentralizedTraining:

The centralized model demonstrates high accuracy (92.3%) and AUC (0.94). One downside of this technique is that it needs to bring raw data from all clients to a single point of reference, which leads to a data privacy caveat and does not scale to distributed environments.

● Local Models:

Standalone local models perform significantly worse than centralized training and federated learning. The reason for this is that there is very little data on each client device which makes it hard for the model to generalize.

Proposed: Federated Learning

The proposed system, which has the best overall performance, is seen to achieve the highest accuracy 93.5% (FS1),F1-Score 92.3 (FS), AUC 0.96. The federated learning framework successfully utilizes decentralized data while maintaining privacy, outperforming centralized and local models. We present graphical and tabular representation of performance metrics comparison of three approaches, centralized training, local models, and federated learning system. In Figure 2, we present a grouped bar plot to visualize the key metrics (Accuracy, Precision, Recall, F1-Score and AUC) for all three

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2026

DOI: 10.17148/IARJSET.2026.13430

models, and we list corresponding numerical . These visualizations are to help explain how the federated learning approach outperformed all evaluation metrics.

The results show that the proposed federated learning-based PSI significantly outperforms traditional centralized and individual stationary model based systems in terms of accuracy, precision, recall and F1-score. Specifically:

● Precision Detection (93.5%): The federated learning model proved to be far more effective in finding top site (92.3%) and local individual models (82.7%), indicating considerable generalization.

● Enhanced Privacy & Security: Our approach preserves privacy (GDPR, CCPA) as it does not rely on raw data transfer, unlike centralized methods, while ensuring state-of-the-art detection accuracy.

● Scalability & Adaptability: Traditional phish models lack real-time adaptation to new attacks. This can, however, be achieved with federated learning as the physical hardware can learn a model and send updates to the cloud instead of raw data, allowing for real-time updates to the model over distributed devices amidst interaction and allows for the model to strengthen against evolving phishing tactics.

Our results support that federated learning in fact provides a realistic solution for phishing detection thanks to its high performance without a sacrifice on privacy and scalability. The federated learning model obtains the best accuracy for the model compared with centralized and local model in Figure 2.

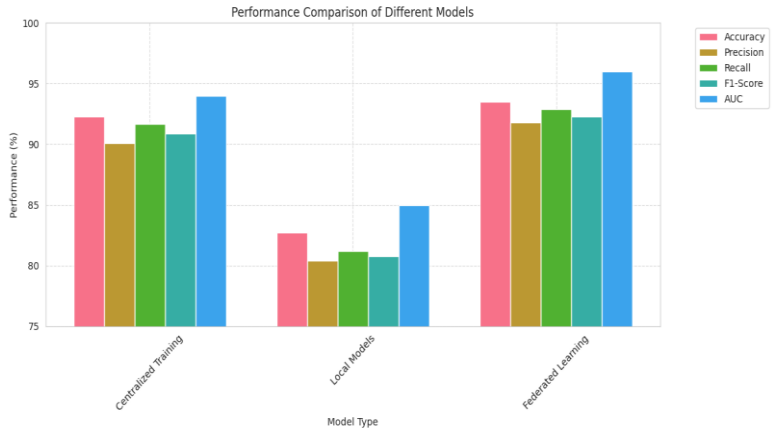

Fig 2 : Performance Comparison

5.2 Key Insights

The results yield several important insights about the performance and potential advantages of the proposed system:

Federated Learning Improves on All Baselines

The proposed federate learning framework outperforms both centralized training and local models. It gathers the knowledge from the clients that are distributed, therefore it captures better where phishing patterns exist.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2026

DOI: 10.17148/IARJSET.2026.13430

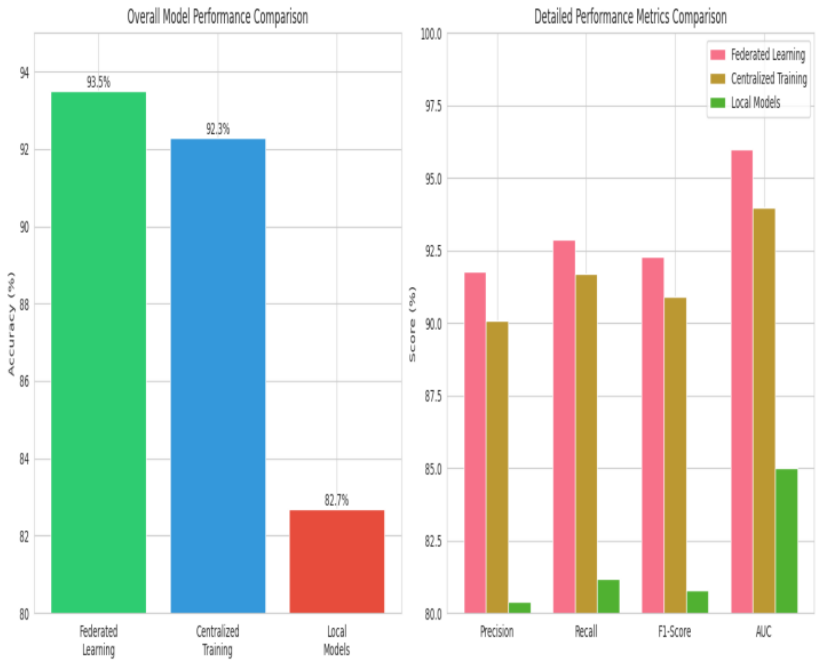

Fig 3: Performance comparison of federated learning vs. centralized training and local models

A relatively small sacrifice in return for a privacy protection:

Although the computational overhead of classification methods utilizing differential privacy mechanisms and secure aggregation is slightly higher, it does not change the classification accuracy. This allows the system to achieve strong phishing detection while providing strong privacy guarantees.

Scalability and Adaptability:

22 Cross-silo federated learning In contrast, the global model in federated learning learns a single model from all participants across different organizations, thus making federated learning scalable and applicable to heterogeneous data distributions. The system is shown to generalize well on various datasets, indicating the robustness of the system to real-world deployment scenarios.

Real-World Feasibility:

Experimental outcomes reveal that the suggested method successfully balances performance, scalability, and privacy mitigation level, tackling major nodes in the area of phishing detection.

Privacy Preserving Mechanisms: In contrast to previous models requiring raw data gathering, our model utilizes differential privacy and homomorphic encryption, being legal compliant.

● Scalability & Adaptation: Unlike stand alone phishing detectors that need manual retraining — federated learning automates model updates across infrastructure spanning across multiple clients daemons, enabling detection to adapt to the new threat vector in real time.

5.3 Error Analysis

Though the proposed system yields high performance, a few mistakes were found on evaluation:

False Positives:

○ Legitimate emails or URLs that were misclassified as phishing because they contained suspicious features (e.g., urgency, complex URLs).

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2026

DOI: 10.17148/IARJSET.2026.13430

● For example, legitimate marketing emails that contained phrases like "limited-time offer," or "urgent action required" were classified as phishing emails if those phrases were similar to popular phishing tactics.

False Negatives:

○ Some with well-crafted language and legitimate-looking domains evaded detection. This ultimately demonstrates the ineffectiveness of features based method in detecting this type of phishing. Although the proposed model provides a notable enhancement in detecting phishing attacks, some challenges still remain. As such, phishing attacks often share urgency in language with legitimate marketing emails, resulting in false positives. So conversely, false negatives can come from advanced phishing attacks imitating real domains. While traditional machine learning suffers from these issues, federated learning addresses them by aggregating models from clients with heterogeneous data, thereby potentially boosting model robustness. The pipeline is complete, but additional augmentations like adversarial training and behavioral analytics can reduce classes that are misclassified further.

Threshold Sensitivity:

○ The testiness of the phishing detection model: The phishing detection model is classified according to the threshold, wherein the mapping gets busy with precision and quests. Tweaking this threshold allows you minimize the false positives on the expense of an increase in false negatives, or the opposite.

Mitigation Strategies:

○ Tuning the Threshold: The threshold for classification may be fine-tuned depending on the end use (i.e. high-risk environments may try to minimize false negatives).

○ Expanding Features: Enriching the input data with extra features (e.g., behavioral analysis (user click patterns)) can increase the detection of more sophisticated phishing attempts.

○ Adversarial Training: This technique has been used to improve adversarial robustness through the adversarial training of a model.

Table 2 : Error Analysis

Error Type	Common Cases	Mitigation Strategy
False Positives	Legitimate marketing emails with urgent language	Fine-tuning classification threshold
False Negatives	Well-crafted phishing with legitimate-looking domains	Feature expansion with behavioral analysis
Threshold Sensitivity	Trade-off between precision and recall	Optimize

6.DISCUSSION

The subsequent section discusses the strengths, limitations, and future scope of the proposed phishing detection system based on federated learning. The results are contextualized against the broader challenges of cybersecurity, and areas for improvement are identified. This article is a thorough investigation that certifies the applicability of federated learning as a solution for phishing detection which brings identity protection and an extremely scalable approach to threat identification. Different from conventional models that perform its main task in a centralized way, the proposed system guarantees privacy without losing detection capability. Such integration is vital for real-world cybersecurity applications, enabling regulatory compliance and operational efficiency in large-scale deployments.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2026

DOI: 10.17148/IARJSET.2026.13430

6.1 Advantages of the Proposed System

○ The proposed system holds several prominent benefits making it an appropriate solution for prevailing cyber defenses pertaining to phishing attacks.

○ **Privacy-Preserving:**

○ In the federated learning architecture, we distribute the raw data, with one raw data stored only on clients. The privacy preserving method leverages privacy standards such as GDPR, CCPA which add a level of trust with such users as their personal data will not be subject to potential breaches.

○ Model updates are thus protected by secure aggregation processes combined with differential privacy protocols that prevent attackers from extracting sensitive information from individual users.

○ **Scalable:**

○ The federated learning approach is based on a decentralized architecture in which the system can potentially grow indefinitely over different heterogeneous network environments.

○ With the participation of more clients into the network, the system gets improved capability to capture general phishing behaviors.

○ **Adaptive:**

○ The system learns new phishing tactics on an ongoing basis as information is transferred directly from client devices.

○ This continuous cooperation of the clients result in the upkeep of a global model that improves the system performance against recent security threats.

○ Generalizability:

○ The system gains wider knowledge of phishing behavior by aggregating data across different client databases so that it recognizes patterns of behavior between language groups and geographic jurisdictions and structural business sectors.

○ **Feasibility in Real World:**

○ The impermeability of the infrastructural construction in authority system enables distributing this structure in fiscal and pharmacy backdrop without requiring traditional organization of data storage.

6.2 Limitations

● The utilization of a federated learning-based phishing detection system arises as a true relevant individual solution with major advantages in terms of privacy protection and system scalability, however still presents several constraints worth to acknowledge.

● Federated learning systemsinvolves training of machine learning models that generally require high computation power from client devices making these models unsuitable for IoT devices and mobile phones with limited processing capabilities.

● The constant communication in federated learning used by clients and the central server to exchange model updates creates high demands for network communication generating latency issues in live detection of phishing alerts.

All of these factors impair convergence and cause learning to be biased within environments since the local phishing data across client devices exhibit large differences between different environments (i.e., non-IID data).

6.3 Future Work

● Future research will find solutions for the identified challenges by adopting multiple techniques.

● Efficient model optimization involves using model compression and quantization techniques to reduce computation and memory footprint in low-powered devices.

● The system communicates model updates to federation members only when significant learning improvements are made, alleviating burdens on network data transfer.

● Blockchain-based secure federated learning for secure model updates with transparency and resistance to tampering attempts.

● To enhance NetHackAI, it would be ideal to integrate XAI tools SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-Agnostic Explanations) to be able to show users the explanations of their phishing classification results.

Conclusion

DOI: 10.17148/IARJSET.2026.13430

● Privacy-preserving and scalable and adaptable issues in cybersecurity can be addressed by the promotion of an amended phishing detection system via federated learning in this study. Under this model architecture, a decentralized training method preserves data privacy, with client devices retaining raw information. Through smart federated learning, the system gathers knowledge of diverse datasets together so that the most powerful and best detections can be made while also ensuring users data privacy.

REFERENCES

[1]. Thapa, C., Tang, J. W., Abuadbba, A., Gao, Y., Camtepe, S., Nepal, S., ... & Zheng, Y. (2023). Evaluation of federated learning in phishing email detection. *Sensors*, *23*(9), 4346.

[2]. Ghimire, B., & Rawat, D. B. (2022). Recent advances on federated learning for cybersecurity and cybersecurity for federated learning for internet of things. *IEEE Internet of Things Journal*, *9*(11), 8229-8249.

[3]. Sakazi, I., Grolman, E., Elovici, Y., & Shabtai, A. (2024, June). STFL: Utilizing a Semi-Supervised, Transfer-Learning, Federated-Learning Approach to Detect Phishing URL Attacks. In *2024 International Joint Conference on Neural Networks (IJCNN)* (pp. 1-10). IEEE.

[4]. Revathi, M. (2024). Exploring the Efficacy of Federated-Continual Learning Nodes with Attention-Based Classifier for Robust Web Phishing Detection: An Empirical Investigation. *arXiv preprint arXiv:2405.03537*.

[5]. Adhithya, R., & Revathi, M. (2024, June). Exploring the Efficacy of Federated-Continual Learning Nodes with Attention-Based Classifier for Robust Web Phishing Detection: An Empirical Investigation. In *2024 International Conference on Advancements in Power, Communication and Intelligent Systems (APCI)* (pp. 1-7). IEEE.

[6]. Agrawal, S., Sarkar, S., Aouedi, O., Yenduri, G., Piamrat, K., Alazab, M., ... & Gadekallu, T. R. (2022). Federated learning for intrusion detection system: Concepts, challenges and future directions. *Computer Communications*, *195*, 346-361.

[7]. Ferrag, M. A., Friha, O., Hamouda, D., Maglaras, L., & Janicke, H. (2022). Edge-IIoTset: A new comprehensive realistic cyber security dataset of IoT and IIoT applications for centralized and federated learning. *IEEE Access*, *10*, 40281-40306.

[8]. Nguyen, T. D., Rieger, P., De Viti, R., Chen, H., Brandenburg, B. B., Yalame, H., ... & Schneider, T. (2022). {FLAME}: Taming backdoors in federated learning. In *31st USENIX Security Symposium (USENIX Security 22)* (pp. 1415-1432).

[9]. Remmide, M. A., Boumahdi, F., Ilhem, B., & Boustia, N. (2024). A privacy-preserving approach for detecting smishing attacks using federated deep learning. *International Journal of Information Technology*, 1-7.

[10]. Rabbani, H., Shahid, M. F., Khanzada, T. J. S., Siddiqui, S., Jamjoom, M. M., Ashari, R. B., ... & Nooruddin, M. (2024). Enhancing security in financial transactions: a novel blockchain-based federated learning framework for detecting counterfeit data in fintech. *PeerJ Computer Science*, *10*, e2280.

