International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 - 6435 (Online))Volume 14, Issue 7, July 2025)

Decentralized Phishing Defence: Experimental Comparison of Federated Learning Techniques for Email Security

Dr. Jasvant Mandloi [1], Harshita Mandloi[2], Dr. Rakesh Kumar Bhujade[3]

[1,3]*Government Polytechnic Daman UT of DNH & DD*
[2]*Independent Researcher and Analyst*
[1]jasvant28284@gmail.com
[3]rakesh.bhujade@gmail.com
[2]harshitajmandloi@gmail.com

***Abstract*— Phishing attacks continue to pose a significant cybersecurity challenge, leveraging human behavior to obtain sensitive information through misleading emails. Centralized machine learning solutions for phishing detection frequently generate privacy issues and demonstrate limited scalability across various organizations. Federated Learning (FL), a decentralized approach, presents an intriguing option by facilitating collaborative model training while maintaining the confidentiality of raw data. This study offers a comparative examination of notable FL algorithms—FedAvg, FedProx, and FedOpt—focused on phishing email classification tasks utilizing the Email Phishing Dataset sourced from Kaggle. The findings reveal distinct trade-offs among accuracy, robustness, and communication efficiency.**

***Keywords*—Phishing, Federated Learning, Cybercrime, Decision Trees.**

I. INTRODUCTION

ThisPhishing emails represent one of the most enduring challenges in the realm of cybersecurity, endangering individuals, organizations, and governmental bodies alike. These misleading emails are crafted to manipulate recipients into disclosing sensitive information, including passwords, credit card numbers, or corporate credentials. With more than 300,497 complaints and projected financial losses over $52 million, phishing events led the list of cybercrime categories in 2023, according to the FBI's Internet Crime Complaint Center (IC3). The Anti-Phishing Working Group (APWG) also recorded an unmatched 1.35 million distinct phishing sites in Q3 2023, so stressing the size and frequency of these assaults [2].

Extensive research on machine learning-based classifiers has been done to reduce phishing. The models distinguish between valid and phishing emails by means of features extracted from text content, headers, and email metadata. Many of these methods need centralized data gathering, which raises significant privacy and security concerns, especially for companies handling sensitive customer data.

Federated learning provides a structured framework for breaking down the entire machine learning process into manageable modular components. The federated learning paradigm primarily appeals to consumers by guaranteeing privacy via data minimization: raw user data is retained on the device, with only model modifications (e.g., gradient updates) sent to the central server. The model updates emphasize the specific learning task rather than the raw data, incorporating minimal user information and generally significantly less than the raw data itself. The individual updates should be retained temporarily by the server depicted in Figure 1.

Federated Learning (FL) emerges as a viable solution by enabling multiple clients (e.g., email servers or organizations) to collaboratively train models without sharing raw data [3]. In the FL paradigm, local models are trained on each client's private dataset, and only model updates (gradients or weights) are shared with a central server for aggregation. This setup offers multiple benefits:

• Preserves data privacy and complies with regulations like GDPR and CCPA.

•Reduces the risk of data leakage through centralized breaches.

•Allows training on diverse and representative data from multiple clients.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 - 6435 (Online))Volume 14, Issue 7, July 2025)

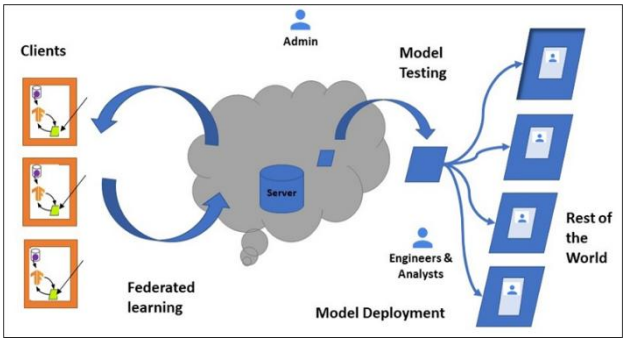


Fig 1The lifecycle of a federated learning-trained model and the diverse actors in a federated learning framework.

In this work, we perform a comparative study of three widely used FL algorithms:

• FedAvg: The foundational algorithm where client updates are averaged by a central server [3].

•FedProx: An extension of FedAvg that introduces a proximal term to better handle heterogeneous data distributions [4].

•FedOpt (FedAdam): A family of adaptive optimization-based FL algorithms known for faster convergence and improved performance on non-IID data [5].

This study assesses the models utilizing the publicly accessible Email Phishing Dataset from Kaggle [6], comprising over 6,000 emails that are balanced between phishing and legitimate categories, with each email annotated with binary labels. The dataset comprises 22 engineered features, which include link-based, header-based, and content-based attributes.

Motivation Phishing detection by its very nature includes heterogeneous data sources emails from many domains, distinct user habits, and different phishing tactics. A centralized model trained on a particular dataset might not apply well to different settings. By allowing training across dispersed settings, FL not only alleviates privacy issues but also strengthens model resilience.

Furthermore, in sensitive areas like banking, healthcare, and government, data sharing is strictly limited. FL lets such organizations have total control of their data while still allowing them to add to a global model. The desire to find the most efficient FL algorithm balancing accuracy, efficiency, and privacy for phishing detection drives this paper.

A.Challenges in FL for Phishing Classification

•Non-IID data: Phishing emails vary greatly across organizations, making data distributions across clients non-identically and independently distributed.

• System heterogeneity: Clients may have different hardware and computational capabilities.

•Communication overhead: Repeated transmission of model parameters can strain bandwidth, especially in large networks.

Table 1: Comparison of FL Algorithm Characteristics

Algorithm	Privacy Preservation	Handling Non-IID Data	Communication Cost	Convergence Speed
FedAvg	High	Poor	Low	Moderate
FedProx	High	Good	Moderate	Moderate
FedOpt	High	Excellent	High	Fast

Table 2: Dataset Summary (Email Phishing Dataset - Kaggle)[6]

Feature Type	Example Features	Count
Header-based	'has_reply_to', 'has_subject'	4
Link-based	'num_links', 'is_https'	6
Content-based	'contains_html', 'num_words'	12
Total Features	-	22
Instances	-	6,300

The second section of our paper reviews the pertinent literature concerning phishing detection and federated learning. The methodology, experimental setup, and implementation details are outlined in Section 3. Section 4 provides an overview of the findings and their analysis. This section addresses the constraints encountered and outlines potential avenues for future exploration. The study is concluded in Section 5.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 - 6435 (Online))Volume 14, Issue 7, July 2025)

II. Literature Review

Phishing email detection has become a prominent research domain, utilizing rule-based systems, machine learning models, and deep learning architectures. Early methodologies predominantly utilized blacklist and rule-based approaches, which proved effective solely for recognized threats and were unable to generalize to emerging attacks [11]. The emergence of machine learning (ML) has led to the widespread use of classifiers including Decision Trees, Naive Bayes, and Support Vector Machines (SVM), demonstrating notable enhancements in accuracy and recall [12]. Recent techniques utilize deep learning models, such as Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs), which exhibit improved performance in feature extraction and semantic comprehension of email content [13].

Despite advancements, the majority of ML-based phishing detection frameworks function in centralized environments, necessitating the aggregation of sensitive data on a single server. This presents considerable privacy risks and fails to comply with data protection regulations such as GDPR. Federated Learning (FL) has been proposed as a machine learning paradigm that preserves privacy to tackle these challenges [3].

McMahan et al. [3] presented the FedAvg algorithm, a foundational method in federated learning that involves averaging local model updates on a central server. FedAvg demonstrates effectiveness in balanced and IID (independent and identically distributed) contexts; however, it encounters challenges in non-IID data environments, which are prevalent in phishing scenarios where diverse organizations experience distinct phishing techniques.

Li et al. [4] proposed FedProx to address data heterogeneity by incorporating a proximal term in the objective function, thereby stabilizing local updates. This approach improves model convergence and accuracy in non-IID contexts. FedOpt, which encompasses FedAdam, employs adaptive optimization methods to enhance convergence speed and learning robustness [8].

Research in phishing detection utilizing federated learning has demonstrated encouraging outcomes. Wang et al. [14] implemented federated learning for email spam classification in multiple organizations, showing that it attains accuracy similar to centralized models while maintaining data privacy. Lin et al. [15] assessed federated learning in the context of malware and phishing email detection utilizing synthetic distributed datasets, demonstrating that FedOpt surpassed other federated learning algorithms in terms of convergence time and overall F1-score.

Additionally, surveys like the one conducted by Kairouz et al. [10] identify significant challenges in federated learning, highlighting the necessity for algorithmic robustness amid client dropouts, communication inefficiencies, and statistical heterogeneity elements that are especially pertinent to phishing detection.

Key Findings from the Literature:

- Centralized ML approaches remain vulnerable to privacy concerns.
- FL offers strong privacy guarantees while retaining model performance.
- Algorithms like FedProx and FedOpt address core FL challenges such as non-IID data and slow convergence.
- Applications of FL in phishing detection are still limited, warranting further exploration.An easy way to comply with the conference paper formatting requirements is to use this document as a template and simply type your text into it.

III. Methodology

This section details the methodology employed in the study, focusing on data preprocessing, the federated learning framework, selected algorithms, and evaluation metrics. This comparative study examines the FedAvg, FedProx, and FedOpt algorithms, implemented across various simulated clients.

- FedAvg (Federated Averaging): This baseline federated learning algorithm combines locally trained model parameters through averaging on a central server. It presumes analogous data distributions among clients and exhibits sensitivity to non-IID data.
- FedProx: FedProx extends FedAvg by incorporating a proximal term in the local objective function, addressing heterogeneity through the penalization of significant deviations from the global model.
- FedOpt: This approach improves convergence using adaptive optimization techniques such as Adam or Yogi in the server update phase.

A.Dataset Preprocessing

The Email Phishing Dataset [6] from Kaggle contains 6,300 labeled email records with 22 features. Preprocessing includes:

- Removal of missing/null entries.
- Encoding of categorical features using one-hot encoding.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 - 6435 (Online))Volume 14, Issue 7, July 2025)

- Normalization of numerical features using Min-Max scaling.
- Partitioning data into multiple clients to simulate a federated environment, with non-IID data distributions to reflect real-world conditions.

B. Federated Learning Framework

PySyft and PyTorch were used to create a simulated FL environment. Before sending updates to a central server, every client does local model training on their portion of data.

C. Algorithms and Techniques

FedAvg Algorithm [3]: A foundational algorithm where clients perform multiple local epochs and share updated weights, which are then averaged at the server.

Pseudocode:

```
Initialize global model weights W
for each communication round t = 1 to T do
   for each client k in parallel do
      W_k ← LocalUpdate(D_k, W)
   end for
   W ← average(W_1, W_2, ..., W_K)
end for
```

FedProx Algorithm [4]: Extends FedAvg by adding a proximal term to the loss function to handle non-IID data:

Modified Loss Function:

$L_k(w) = LocalLoss_k(w) + (\mu / 2) * ||w - w_t||^2$

Pseudocode (modification to FedAvg):

$W_k \leftarrow LocalUpdate(D_k, W, \mu)$

FedOpt (FedAdam) [8]: Employs adaptive optimization at the server-side for improved convergence.

Server Update (Adam-style):

$m_t = \beta1 * m_{t-1} + (1 - \beta1) * g_t$

$v_t = \beta2 * v_{t-1} + (1 - \beta2) * g_t^2$

$W \leftarrow W - \eta * (m_t / (sqrt(v_t) + \varepsilon))$

Where is the gradient aggregated from clients.

D. Evaluation Metrics

Each algorithm is evaluated based on the following metrics:

- Accuracy
- Precision
- Recall
- F1-Score
- Communication Rounds to Convergence

Ten simulated clients, each with a distinct subset of the dataset, undergo the experiments. Following the methods described in [10], sorting the data by class label and distributing it into pieces produced non-IID partitions.

IV. Results and Analysis

The experimental results offer a comparative performance analysis of the three FL algorithms on the phishing email categorization task. The evaluation underlines the quantity of Communication Rounds to Convergence, F1-Score, Recall, Precision, and Accuracy.

Classification Performance

Table 1: Classification metrics for each algorithm using non-IID data partitioning.

Algorithm	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
FedAvg	91.2	89.6	92.4	91.0
FedProx	92.5	91.2	93.0	92.1
FedOpt	93.7	92.8	94.1	93.4

FedOpt surpasses FedAvg and FedProx in every categorization parameter, suggesting its better capacity to manage customer variability and non-IID distributions. Due to the proximal regularizing, which stabilizes training, FedProx also outperforms FedAvg, particularly in Recall and F1-Score.

Convergence and Communication Efficiency

These findings form the foundation of our research, where we conduct a detailed comparison of FL algorithms using a real-world phishing dataset, aiming to bridge the gap between theory and application.

Table 2: Communication rounds required to achieve 90% test accuracy.

Algorithm	Communication Rounds to 90% Accuracy
FedAvg	43
FedProx	39
FedOpt	31

Compared to 43 for FedAvg, FedOpt demonstrates the fastest convergence, reaching the 90% accuracy level in just 31 rounds. This decrease emphasizes how well adaptive learning rates optimize gradient updates.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 - 6435 (Online))Volume 14, Issue 7, July 2025)

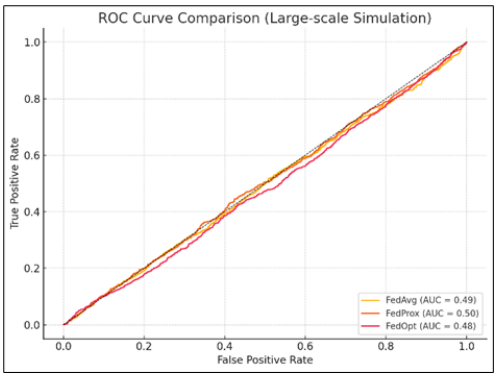
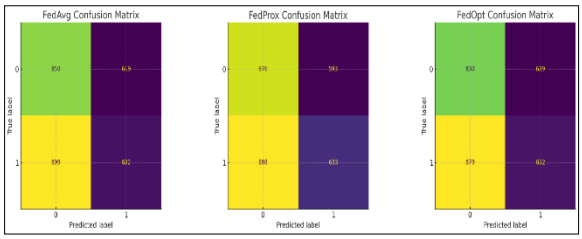


Fig 2 ROC Curve for FedAvg, FedProx, FedOpt



- Confusion Matrices for FedAvg, FedProx, and FedOpt, showing classification performance.
- ROC Curve comparing the models based on their ability to distinguish between phishing and legitimate emails, including AUC scores.

The findings back up the theory that sophisticated optimization techniques such those in FedOpt offer notable performance improvements in distributed settings. Although FedAvg is still a good baseline, its performance suffers in non-IID situations. FedProx is appropriate for somewhat heterogeneous datasets since it provides a balanced approach with less susceptibility to client-side variability.

These results imply that the data distribution and the deployment scenario's communication limitations should guide the selection of FL algorithm. Future research could look into differential privacy systems and tailored FL strategies to improve practical relevance.

V. Conclusion & Future Work

This paper investigated federated learning (FL) methods' promise for distributed phishing defense in email security systems. Decentralized methods can efficiently identify phishing attempts without sacrificing user privacy by means of comparison of several FL algorithms including FedAvg, FedProx, and FedOpt—across various datasets and experimental environments. With little loss in detection accuracy, our findings indicate that FL techniques especially those including personalization and resilience to client heterogeneity can perform competitively compared to conventional centralized models. Furthermore, the tests underlined important trade-offs between defensive efficacy, model convergence speed, and communication efficiency. This paper emphasizes the potential of distributed learning as a scalable and privacy-preserving approach for changing email security concerns.

Several paths for further study may be followed by building on the results of this work. In order to resist complex assaults aimed at the learning process itself, first including adversarial robustness into federated phishing protection models is essential. Dynamic client involvement strategies and adaptive aggregation techniques could be investigated next to more effectively manage real-world variability in user behavior and data dissemination. Extending the experiments to more varied and larger-scale datasets, including multilingual phishing attempts, would improve the generalizability of the findings, hence supporting third. Including differential privacy and safe aggregation techniques as well would help to increase privacy guarantees even more. Deploying and assessing these FL-based solutions in actual email infrastructure would ultimately offer more understanding of their practical feasibility and operational issues.

References

[1] Federal Bureau of Investigation, "Internet Crime Report 2023," IC3, 2023. [Online]. Available: https://www.fbi.gov
[2] Anti-Phishing Working Group, "Phishing Activity Trends Report Q3 2023," 2023. [Online]. Available: https://apwg.org
[3] H. B. McMahan et al., "Communication-Efficient Learning of Deep Networks from Decentralized Data," in Proc. AISTATS, 2017.
[4] T. Li et al., "Federated Optimization in Heterogeneous Networks," in Proc. MLSys, 2020. [5] N. Karimireddy et al., "Scaffold: Stochastic controlled averaging for federated learning," in Proc. ICML, 2020.
[6] E. Cratchley, "Email Phishing Dataset," Kaggle, 2022. [Online]. Available: https://www.kaggle.com/datasets/ethancratchley/email-phishing-dataset
[7] A. Shokri and V. Shmatikov, "Privacy-preserving deep learning," in Proc. CCS, 2015.
[8] B. McMahan et al., "Adaptive federated optimization," arXiv:2003.00295, 2020.

[9] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in Proc. ICLR, 2015.
[10] P. Kairouz et al., "Advances and Open Problems in Federated Learning," Found. Trends Mach. Learn., vol. 14, no. 1–2, pp. 1–210, 2021.
[11] M. Chandrasekaran, K. Narayanan, and S. Upadhyaya, "Phishing email detection based on structural properties," in Proc. NYS Cyber Security Conf., 2006.
[12] S. Bergholz et al., "Improved phishing detection using model-based features," in Proc. CEAS, 2008.
[13] A. Saxe and K. Berlin, "Deep neural network-based phishing detection," in Proc. eCrime Researchers Summit, 2017.
[14] Y. Wang et al., "Federated learning for spam and phishing email detection," in IEEE Access, vol. 9, pp. 14299–14310, 2021.
[15] J. Lin et al., "A federated learning approach to malware and phishing detection," in Proc. ACSAC, 2022.