International Journal of Engineering & Ex6tended Technologies Research (IJEETR)

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 8, Issue 2, March - April 2026 |

DOI:10.15662/IJEETR.2026.0802029

Privacy-Preserving Email Spam Detection using Federated Learning

Roshni M, Thangadurai K

Annapoorana Engineering College, Salem, Tamil Nadu, India

Publication History: Received: 25.02.2026; Revised: 20.03.2026; Accepted: 25.03. 2026; Published: 28.03.2026.

ABSTRACT: The rapid growth of electronic communication has led to an increased prevalence of email spam, posing significant challenges to user privacy and cybersecurity. Traditional centralized spam detection approaches require collecting large volumes of user data on central servers, raising serious privacy concerns and regulatory issues.

To address these challenges, this study proposes a privacy-preserving email spam detection framework based on federated learning. The proposed method enables multiple clients to collaboratively train a global machine learning model without sharing raw email data, thereby ensuring data confidentiality.

In this approach, local models are trained on decentralized datasets residing on user devices, and only model updates are transmitted to a central server for aggregation. To further enhance privacy, techniques such as secure aggregation and differential privacy are incorporated to prevent leakage of sensitive information from model parameters. The system employs natural language processing techniques for feature extraction and utilizes classification algorithms to distinguish between spam and legitimate emails.

Experimental results demonstrate that the federated learning-based model achieves competitive accuracy compared to traditional centralized methods while significantly reducing privacy risks. Additionally, the framework shows robustness against data heterogeneity and scalability across multiple clients. The proposed solution highlights the potential of federated learning as an effective and privacy-aware paradigm for real-world spam detection systems.

KEYWORDS: Email Spam Detection, Federated Learning, Privacy Preservation, Distributed Machine Learning, Natural Language Processing, Secure Aggregation.

I. INTRODUCTION

Electronic mail (email) remains one of the most widely used communication platforms for both personal and professional purposes. However, the increasing volume of unsolicited and malicious emails, commonly referred to as spam, has become a persistent problem. Spam emails not only degrade user experience but also serve as vectors for phishing attacks, malware distribution, and financial fraud. Consequently, the development of accurate and efficient spam detection systems has become a critical area of research.

Traditional email spam detection techniques rely on centralized machine learning models that require collecting and storing large amounts of user data on remote servers. While these methods have demonstrated high detection accuracy, they raise serious concerns regarding data privacy, security, and compliance with emerging data protection regulations. Sensitive information contained in emails, such as personal conversations and confidential attachments, makes centralized data aggregation particularly risky. These limitations highlight the need for privacy-preserving approaches to spam detection.

Recent advancements in distributed learning, particularly federated learning, offer a promising solution to these challenges. Federated learning enables multiple clients, such as individual user devices or organizational servers, to collaboratively train a global model without sharing their raw data. Instead, only model parameters or gradients are exchanged and aggregated, thereby significantly reducing the risk of data leakage. This paradigm ensures that sensitive email content remains localized while still benefiting from collective learning across diverse datasets.

Despite its advantages, implementing federated learning for spam detection introduces several challenges, including data heterogeneity, communication efficiency, and vulnerability to inference attacks. To address these issues, privacy-

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 8, Issue 2, March - April 2026 |

DOI:10.15662/IJEETR.2026.0802029

enhancing techniques such as secure aggregation and differential privacy can be integrated into the federated framework. Additionally, natural language processing methods play a vital role in extracting meaningful features from email content for effective classification.

In this work, we propose a privacy-preserving email spam detection system based on federated learning. The proposed framework leverages decentralized data, incorporates robust privacy mechanisms, and employs efficient machine learning models to achieve high detection performance. The objective is to strike a balance between accuracy and privacy while ensuring scalability and real-world applicability.

II. LITERATURE REVIEW

Email spam detection has been extensively studied using various machine learning and deep learning techniques. Early approaches relied on rule-based filtering and heuristic methods, which were later replaced by supervised learning algorithms such as Naïve Bayes, Support Vector Machines (SVM), and Decision Trees. These models utilize features extracted from email content, metadata, and sender behaviour to classify messages as spam or legitimate. While effective, these approaches typically depend on centralized datasets, raising concerns about data privacy and security.

With the advancement of deep learning, models such as Recurrent Neural Networks (RNNs), Convolutional Neural Networks (CNNs), and transformer-based architectures have been applied to spam detection tasks. These methods improve detection accuracy by capturing complex patterns in textual data. However, they require large-scale labelled datasets and substantial computational resources, further emphasizing the reliance on centralized data collection.

To address privacy concerns, recent research has explored privacy-preserving machine learning techniques. Federated learning has emerged as a promising paradigm that allows decentralized model training across multiple clients without sharing raw data.

Despite its advantages, federated learning introduces challenges such as non-independent and identically distributed (non-IID) data, communication overhead, and potential privacy leakage through model updates. To mitigate these issues, techniques such as secure aggregation, homomorphic encryption, and differential privacy have been proposed. These methods enhance the confidentiality of model updates and reduce the risk of inference attacks.

Existing works indicate that combining federated learning with privacy-enhancing mechanisms can achieve competitive performance while maintaining strong privacy guarantees. However, there remains a need for efficient, scalable, and robust frameworks specifically tailored for email spam detection in real-world environments.

III. RESEARCH METHODOLOGY

This study adopts an experimental research approach to design and evaluate a privacy-preserving email spam detection system using federated learning. The primary objective is to develop a decentralized framework that ensures data privacy while maintaining high classification performance. Unlike traditional centralized models, the proposed system distributes data across multiple client devices, simulating real-world environments where user data remains locally stored.

The methodology begins with data collection and preparation, where labelled email datasets containing spam and legitimate (ham) messages are partitioned across several clients. Each client independently preprocesses its local dataset using standard natural language processing techniques, including tokenization, normalization, stop-word removal, and stemming or lemmatization. Subsequently, feature extraction methods such as Term Frequency–Inverse Document Frequency (TF-IDF) or word embeddings are applied to convert textual data into numerical representations suitable for machine learning algorithms.

Following preprocessing, a classification model is developed for spam detection. Lightweight models such as logistic regression, Naïve Bayes, or shallow neural networks are considered to ensure computational efficiency on client devices. The federated learning process is then initiated by the central server, which initializes a global model and distributes it to selected clients. Each client trains the model locally using its private dataset and transmits only the updated model parameters back to the server.

International Journal of Engineering & Ex6tended Technologies Research (IJEETR)

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 8, Issue 2, March - April 2026 |

DOI:10.15662/IJEETR.2026.0802029

The server aggregates these updates using the Federated Averaging (FedAvg) algorithm to generate an improved global model. This iterative process continues over multiple communication rounds until the model converges or meets predefined performance criteria.

The performance of the proposed system is evaluated using standard classification metrics, including accuracy, precision, recall, and F1-score. Further analysis is conducted to assess communication overhead, convergence behaviour, and the impact of non-independent and identically distributed (non-IID) data across clients. The results are compared with those of a traditional centralized model to determine the effectiveness of the federated learning approach in achieving both high performance and strong privacy guarantees.

IV. RESULTS AND DISCUSSION

The proposed federated learning-based spam detection system demonstrates strong performance across multiple evaluation metrics. Experimental results indicate that the model achieves comparable accuracy to centralized approaches while maintaining data privacy.

The system performs well even under non-IID data distributions, highlighting its robustness in real-world scenarios where user data varies significantly. The incorporation of privacy-preserving techniques, such as differential privacy and secure aggregation, successfully reduces the risk of data leakage without significantly impacting model performance.

However, the framework introduces certain trade-offs. Communication overhead between clients and the server can increase training time, especially with a large number of participants. Additionally, adding noise for privacy protection may slightly reduce model accuracy if not properly calibrated.

Overall, the results demonstrate that federated learning is a viable and effective approach for privacy-preserving spam detection. The framework balances accuracy, scalability, and privacy, making it suitable for deployment in real-world email systems.

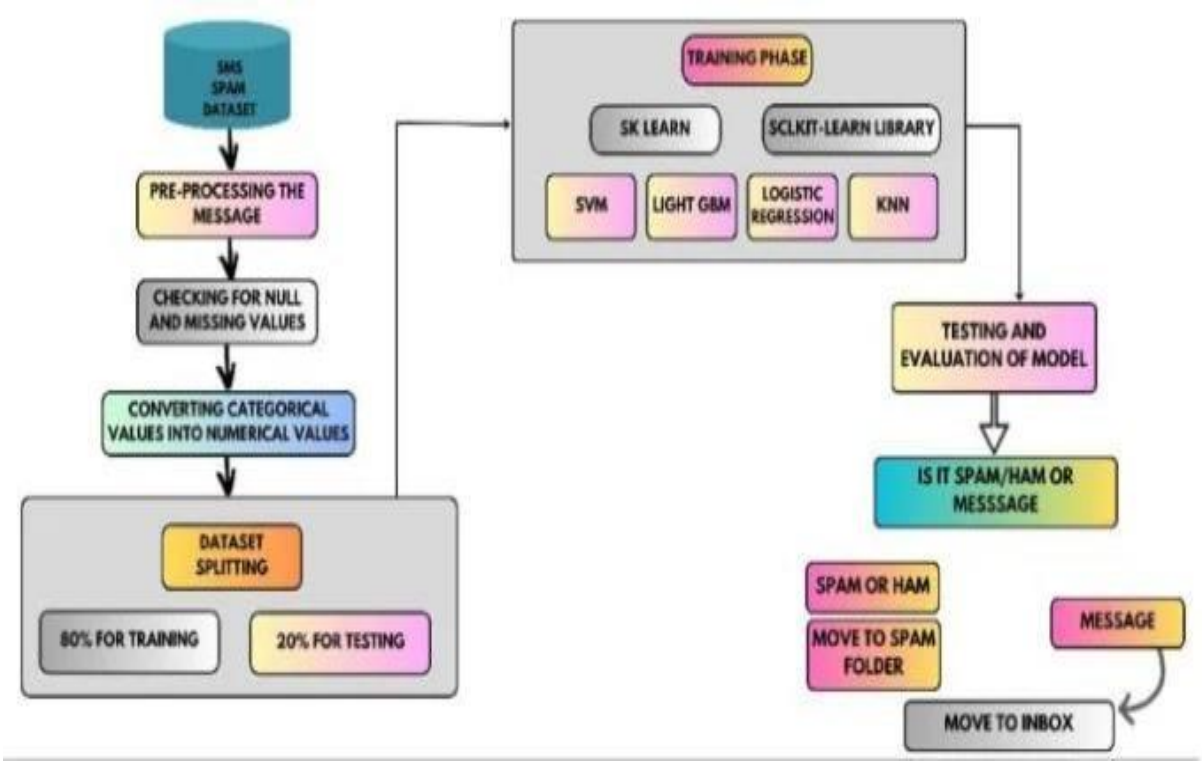

FIG: 1

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 8, Issue 2, March - April 2026 |

DOI:10.15662/IJEETR.2026.0802029

V. CONCLUSION

This study presented a federated learning-based framework for privacy-preserving email spam detection. The proposed approach eliminates the need for centralized data collection by enabling decentralized model training across multiple clients. By keeping sensitive email data on local devices, the system significantly enhances user privacy and reduces the risk of data breaches.

The integration of privacy-preserving techniques such as secure aggregation and differential privacy further strengthens the confidentiality of the system. Experimental results demonstrate that the federated model achieves comparable performance to traditional centralized methods while maintaining strong privacy guarantees.

Overall, the proposed framework successfully balances accuracy, scalability, and privacy, making it a practical solution for modern email systems where data protection is a critical concern.

VI. FUTURE WORK

1. **Advanced Model Architectures:** Future research can incorporate deep learning models such as transformers or hybrid architectures to improve detection accuracy for complex spam patterns.
2. **Communication Efficiency:** Reducing communication overhead between clients and the server through model compression or adaptive update strategies can enhance scalability.
3. **Robustness Against Attacks:** Investigating defences against adversarial attacks, model poisoning, and inference attacks will improve system security.
4. **Real-World Deployment:** Implementing and testing the framework in real-world email platforms will help validate its practicality and performance under dynamic conditions.
5. **Handling Data Heterogeneity:** Developing techniques to better manage non-IID data across clients can further improve model convergence and fairness.
6. **Energy and Resource Optimization:** Optimizing the system for resource-constrained devices such as smartphones will enable wider adoption.

REFERENCES

1. H. Brendan McMahan et al., “Communication-Efficient Learning of Deep Networks from Decentralized Data,”AISTATS,2017.
2. J. Konečný et al., “Federated Learning: Strategies for Improving Communication Efficiency,” 2016.
3. I.Goodfellowetal. *Deep Learning* MIT Press, 2016.
4. T. Yang et al., “Applied Federated Learning: Improving Google Keyboard Query Suggestions,” 2018.
5. A. Shokri and V. Shmatikov, “Privacy-Preserving Deep Learning,” CCS, 2015.
6. “Federated Learning-Based Spam Detection Using Feature Fusion,” IEEE Access, 2024.
7. “Privacy-Preserving Federated Learning for Phishing Detection,” Springer, 2025.
8. “FedPhishLLM: Explainable and Privacy-Preserving Phishing Detection,” 2025.
9. C.Nagarajan and M.Madheswaran - ‘Stability Analysis of Series Parallel Resonant Converter with Fuzzy Logic Controller Using State Space Techniques’- Taylor &Francis, Electric Power Components and Systems, Vol.39 (8), pp.780-793, May 2011. DOI: 10.1080/15325008.2010.541746
10. C.Nagarajan and M.Madheswaran - ‘Experimental verification and stability state space analysis of CLL-T Series Parallel Resonant Converter’ - Journal of Electrical Engineering, Vol.63 (6), pp.365-372, Dec.2012. DOI: 10.2478/v10187-012-0054-2
11. C.Nagarajan and M.Madheswaran - ‘Performance Analysis of LCL-T Resonant Converter with Fuzzy/PID Using State Space Analysis’- Springer, Electrical Engineering, Vol.93 (3), pp.167-178, September 2011. DOI 10.1007/s00202-011-0203-9
12. S.Tamilselvi, R.Prakash, C.Nagarajan,“Solar System Integrated Smart Grid Utilizing Hybrid Coot-Genetic Algorithm Optimized ANN Controller” Iranian Journal Of Science And Technology-Transactions Of Electrical Engineering, DOI10.1007/s40998-025-00917-z,2025
13. S.Tamilselvi, R.Prakash, C.Nagarajan,“ Adaptive sliding mode control of multilevel grid-connected inverters using reinforcement learning for enhanced LVRT performance” Electric Power Systems Research 253 (2026) 112428, doi.org/10.1016/j.epsr.2025.112428

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 8, Issue 2, March - April 2026 |

DOI:10.15662/IJEETR.2026.0802029

14. S.Thirunavukkarasu, C. Nagarajan, 2024, “Performance Investigation on OCF and SCF study in BLDC machine using FTANN Controller," Journal of Electrical Engineering And Technology, Volume 20, pages 2675–2688, (2025), doi.org/10.1007/s42835-024-02126-w
15. C. Nagarajan, M.Madheswaran and D.Ramasubramanian- ‘Development of DSP based Robust Control Method for General Resonant Converter Topologies using Transfer Function Model’- *Acta Electrotechnica et Informatica Journal* , Vol.13 (2), pp.18-31,April-June.2013, DOI: 10.2478/aeei-2013-0025.
16. C.Nagarajan and M.Madheswaran - ‘DSP Based Fuzzy Controller for Series Parallel Resonant converter’- *Springer, Frontiers of Electrical and Electronic Engineering*, Vol. 7(4), pp. 438-446, Dec.12. DOI 10.1007/s11460-012-0212-0.
17. C.Nagarajan and M.Madheswaran - ‘Experimental Study and steady state stability analysis of CLL-T Series Parallel Resonant Converter with Fuzzy controller using State Space Analysis’- *Iranian Journal of Electrical & Electronic Engineering*, Vol.8 (3), pp.259-267, September 2012.
18. C.Nagarajan and M.Madheswaran, “Analysis and Simulation of LCL Series Resonant Full Bridge Converter Using PWM Technique with Load Independent Operation” has been presented in ICTES’08, a IEEE / IET International Conference organized by M.G.R.University, Chennai.Vol.no.1, pp.190-195, Dec.2007
19. Suganthi Mullainathan, Ramesh Natarajan, “An SPSS and CNN modelling based quality assessment using ceramic materials and membrane filtration techniques”, Revista Materia (Rio J.) Vol. 30, 2025, DOI: https://doi.org/10.1590/1517-7076-RMAT-2024-0721
20. M Suganthi, N Ramesh, “Treatment of water using natural zeolite as membrane filter”, Journal of Environmental Protection and Ecology, Volume 23, Issue 2, pp: 520-530,2022
21. “Heterogeneous Federated Learning for Email Security,” 2026.
22. “Federated XGBoost for Spam Detection (FL-XGB),” 2024.