InSight Bulletin: A Multidisciplinary Interlink International Research Journal
Peer Reviewed International, Open Access Journal.
ISSN : 3065-7857 / Website: https://ibrj.us / Volume-3, Issue-1 / January - 2026

Original Article

Privacy-Focused Machine Learning Models for Cybersecurity

Radhika[1], Nagnath Bhiste[2]
[1]Asst. Prof., Department of Computer Science, Dr. D. Y. Patil Arts, Commerce and Science College Akurdi, Pune
[2]Department of Computer Science, Dr. D. Y. Patil Arts, Commerce and Science College Akurdi, Pune

Manuscript ID:
IBMIIRJ -2026-030121

Submitted: 08 Dec. 2025

Revised: 12 Dec. 2025

Accepted: 07 Jan. 2026

Published: 31 Jan. 2026

ISSN: 3065-7857

Volume-3

Issue-1

Pp. 112-115

January 2026

Correspondence Address:
Radhika1, Asst. Prof., Department ofComputer Science, Dr. D. Y. Patil Arts, Commerce and Science College Akurdi, Pune
Email: radhikabhiste19@gmail.com

OPEN ACCESS

Quick Response Code:

Web. https://ibrj.us

DOI: 10.5281/zenodo.18950606

DOI Link:
https://doi.org/10.5281/zenodo.18950606

Creative Commons

Abstract

The rapid adoption of machine learning (ML) in cybersecurity has significantly improved threat detection and response capabilities. However, conventional ML-based security systems often require centralized data collection, leading to serious privacy risks such as data leakage, unauthorized access, and regulatory non-compliance. This paper explores privacy-focused machine learning models for cybersecurity, emphasizing techniques that ensure data confidentiality while maintaining detection accuracy. Approaches such as federated learning, differential privacy, homomorphic encryption, and secure multi-party computation are analyzed in the context of intrusion detection, malware classification, and anomaly detection. The study highlights the trade-offs between privacy, performance, and computational overhead, and discusses real-world challenges in deploying privacy-preserving ML systems. The findings suggest that privacy-aware ML frameworks are essential for building trustworthy, scalable, and regulation-compliant cybersecurity solutions.

Keywords: *Privacy-Preserving Machine Learning, Cybersecurity, Federated Learning, Differential Privacy, Secure AI*

Introduction

Artificial Intelligence (AI) and Machine Learning (ML) have become integral to modern cybersecurity systems by enabling automated threat detection, intrusion prevention, and anomaly analysis. These techniques offer improved accuracy and scalability compared to traditional security mechanisms. However, most AI-based cybersecurity solutions rely on centralized data collection, which raises significant privacy concerns, including data leakage, unauthorized access, and misuse of sensitive information. With the enforcement of data protection regulations such as the General Data Protection Regulation (GDPR), Digital Personal Data Protection (DPDP) Act of India, and Health Insurance Portability and Accountability Act (HIPAA), ensuring data privacy has become a legal and ethical requirement. Consequently, there is a stron need for privacy-focused machine learning models that can effectively defend against cyber threats while preserving data confidentiality. This research aims to analyze privacy-preserving machine learning techniques in cybersecurity and evaluate their role in achieving secure, trustworthy, and regulation-compliant cyber defense systems.

Problem Statement

Traditional machine learning–based cybersecurity systems rely on centralized data collection and processing, which exposes sensitive information to privacy risks such as data breaches, unauthorized access, and regulatory non-compliance. While these systems are effective in detecting cyber threats, they often violate data protection requirements mandated by regulations such as GDPR, DPDP Act (India), and HIPAA. Therefore, there is a critical need to design and evaluate privacy-focused machine learning models that can provide effective cyber threat detection while ensuring data confidentiality and compliance with data protection laws.

Objectives of the Study

1. **To Analyze Privacy Risks in Traditional ML-Based Cybersecurity Systems:**
   Investigate the vulnerabilities associated with centralized machine learning models, including potential data breaches, unauthorized access, and non-compliance with data protection regulations such as GDPR, DPDP Act (India), and HIPAA.

How to cite this article:
Radhika, & Bhiste, N. (2026). Privacy-Focused Machine Learning Models for Cybersecurity. InSight Bulletin: A Multidisciplinary Interlink International Research Journal, 3(1), 112–115. https://doi.org/10.5281/zenodo.18950606

This objective aims to establish the motivation for adopting privacy-preserving methods in cybersecurity applications.

2. **To Study Privacy-Preserving Machine Learning Techniques:**
   Examine various privacy-focused ML methods such as federated learning, differential privacy, homomorphic encryption, and secure multi-party computation. The study will analyze their principles, mechanisms, and applicability in securing sensitive cybersecurity data while maintaining effective threat detection.
3. **To Evaluate the Effectiveness of Privacy-Focused Models in Cyber Threat Detection:**
   Assess how privacy-preserving ML models perform in detecting cyber threats, including intrusions, malware, and abnormal network behaviors. Metrics such as accuracy, precision, recall, F1-score, privacy overhead, and communication overhead will be used for evaluation.
4. **To Identify Trade-offs Between Privacy, Accuracy, and Efficiency:**
   Investigate how implementing privacy-preserving techniques impacts model performance, system efficiency, and resource consumption. This objective helps in understanding the practical limitations and challenges of deploying such models in real-world cybersecurity scenarios.
5. **To Explore Regulatory Compliance and Ethical Considerations:**
   Analyze how privacy-focused ML models can adhere to legal frameworks and ethical standards. The objective is to ensure that models not only secure data but also comply with global regulations and promote responsible AI usage in cybersecurity.
6. **To Recommend Best Practices and Future Directions:**
   Provide insights and recommendations for designing, implementing, and scaling privacy-focused ML systems for cybersecurity. This includes identifying research gaps, suggesting hybrid approaches, and highlighting potential innovations such as integrating explainable AI with privacy-preserving models.

Literature Review (Key Areas)

Technique	Application	Key Findings	Limitation	Sample References
Centralized Machine Learning	Intrusion Detection, Malware Detection	High detection accuracy using ML/DL models	Privacy leakage due to centralized data storage	Sommer& Paxson (2010), Buczak & Guven (2016)
Federated Learning	Distributed IDS, IoT Security	Preserves ata Privacy by decentralized training	High communication overhead, model poisoning risk	Kairouz et al. (2021)
Differential Privacy	Log Analysis, User Behavior Monitoring	Protects individual data through noise injection	Reduced accuracy and utility loss	Dwork (2006)
Encrypted Machine Learning	Secure Data Analytics	Enables computationon encrypted data	High computational and time cost	Shokri Shmatikov (2015)

Privacy-Focused Machine Learning Techniques

Privacy-focused machine learning (ML) techniques aim to enhance cybersecurity systems while preserving the confidentiality of sensitive data. Unlike traditional centralized ML models, these approaches reduce privacy risks by minimizing or eliminating the need to share raw data. Key techniques include:

- **Federated Learning (FL)**

Federated Learning is a decentralized approach where multiple nodes or clients train local models on their private data. Only model updates or gradients are shared with a central server, which aggregates them to create a global model. This ensures that raw data never leaves the local environment. FL is widely applied in intrusion detection systems, malware detection, and IoT security. Its main advantages are enhanced privacy and reduced data transfer, although challenges include communication overhead and susceptibility to adversarial attacks.

- **Differential Privacy (DP)**

Differential Privacy is a mathematical framework that protects individual data points by adding controlled noise to the data or model outputs. This prevents adversaries from inferring sensitive information while still enabling model training. DP can be applied to log analysis, user behavior monitoring, and anomaly detection. However, excessive noise can reduce model accuracy, so a balance between privacy and utility must be maintained.

- **Homomorphic Encryption (HE)**

Homomorphic Encryption allows computation to be performed directly on encrypted data without decryption. This technique provides strong privacy guarantees, making it ideal for secure analytics in sensitive environments. For example, cybersecurity models can analyze encrypted network traffic or system logs without ever accessing raw data. The main limitation is its high computational cost, which can make real-time applications challenging.

- **Secure Multi-Party Computation (SMPC)**

SMPC enables multiple parties to jointly compute a function over their private inputs without revealing them to each other. In cybersecurity, SMPC can facilitate collaborative threat detection across organizations while maintaining data confidentiality. It ensures that each participant contributes to the learning process without disclosing sensitive information, but the approach may

introduce latency and complexity when scaling to many participants.

- **Hybrid Approaches**

Recent research explores combining multiple privacy-preserving techniques, such as federated learning with differential privacy or homomorphic encryption, to achieve enhanced security and privacy simultaneously. These hybrid approaches aim to balance accuracy, efficiency, and privacy, addressing the limitations of individual methods while ensuring compliance with regulatory frameworks.

Proposed Methodology

The proposed methodology employs a privacy-focused machine learning framework to enhance cybersecurity systems while maintaining data confidentiality. The approach leverages decentralized learning, ensuring that sensitive data remains local and is not transmitted to a central server.

- **Data Collection**

Network traffic, system logs, and other cybersecurity-related data are collected at distributed nodes, such as organizational endpoints, IoT devices, or servers. This ensures that private information is kept on local devices, aligning with privacy regulations and reducing the risk of data breaches.

- **Local Model Training**

Each node trains its local machine learning or deep learning model independently using algorithms like Random Forest, Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), and Autoencoders. These models learn to distinguish between normal and malicious behaviors, capturing patterns specific to local environments.

- **Secure Parameter Aggregation**

Instead of sharing raw data, nodes transmit only model updates or parameters to a central aggregation server. Using secure aggregation techniques, the server combines these parameters to generate a global model while ensuring that no sensitive data is exposed. This step preserves privacy and mitigates risks associated with centralized data collection.

- **Global Model Update**

The updated global model is redistributed to all participating nodes, where it can be further refined with local data. This iterative process improves model performance over multiple training rounds while maintaining privacy.

- **Threat Detection Evaluation**

The final global model is evaluated for cyber threat detection performance using standard benchmark datasets such as NSL-KDD, CICIDS 2017, and UNSW-NB15. Evaluation metrics include accuracy, precision, recall, F1-score, privacy overhead, and Communication overhead, allowing comprehensive assessment of the model's effectiveness, efficiency, and privacy preservation.

Figure: Architecture of the Proposed Federated Cyber Threat Detection System

Challenges & Limitations

Privacy-focused machine learning in cybersecurity faces several challenges. Protecting sensitive data can reduce detection accuracy due to techniques like differential privacy. High computational and communication overhead, especially in federated learning and encrypted computation, can affect performance and scalability. Models may also be vulnerable to attacks such as model poisoning, and integrating them with existing systems often requires significant changes. Additionally, compliance with regulations like GDPR, DPDP Act, and HIPAA adds complexity. Despite these challenges, privacy-preserving ML remains essential for secure and trustworthy cyber defense.

Future Scope

Future research in privacy-focused machine learning for cybersecurity includes integrating explainable AI for transparency, developing hybrid federated + XAI models for improved privacy and trust, creating legal-compliant AI security systems, and incorporating AI governance and ethics to ensure responsible and accountable deployment.

Conclusion

Privacy-focused machine learning has emerged as a vital approach in cybersecurity, offering robust threat detection while safeguarding sensitive data and ensuring regulatory compliance. By leveraging techniques such as federated learning, differential privacy, and encrypted computation, organizations can maintain security without compromising privacy. Future research should focus on optimizing the balance between accuracy, efficiency, and privacy to enable scalable and practical deployment of these models in real-world cybersecurity systems.

Acknowledgement

I would like to express my sincere gratitude to Asst. Prof. Radhika Nagnath Bhiste for her valuable guidance, constant encouragement, and insightful suggestions throughout the course of this research work. Her expertise and support played a crucial role in shaping this tudy. I am also thankful to the Computer Science Department, Dr. D. Y. Patil Arts, Commerce and Science College, Akurdi, Pune, for providing the necessary facilities and academic environment required to carry out this research successfully. I extend my heartfelt thanks to all faculty members, friends, and colleagues who directly or indirectly supported and motivated me during the completion of this work. Finally, I express my deep gratitude to my family for their continuous encouragement, patience, and moral support, without which this work would not have been possible

Financial support and sponsorship

Nil.

Conflicts of interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

1. S. A. Mahmud, N. Islam, Z. Islam, Z. Rahman, and S. T. Mehedi, "Privacy-Preserving Federated Learning-Based Intrusion Detection Technique for Cyber-Physical Systems," Mathematics, vol. 12, no. 20, p. 3194, 2024. MDPI
2. "Federated Learning for Cybersecurity: A Privacy-Preserving Approach," Applied Sciences, vol. 15, no. 12, p. 6878, 2025. MDPI
3. T. Asagunla, "Federated Cyber Defense: A Privacy-Preserving AI Framework for Threat Intelligence Sharing Across Multinational Enterprises," Int. J. Sci. Res. Mod. Technol., vol. 2, no. 8, pp. 26–30, 2023. IJSMT
4. E. Shalabi, W. Khedr, E. Rushdy, and A. Salah, "Privacy-Preserving Federated Learning in Network Intrusion Detection: A Systematic Literature Review," Int. J. Computers and Informatics, vol. 8, pp. 23–43, 2025. IJCI
5. "Privacy-Preserving Machine Learning in Cybersecurity," Int. J. Inf. Technol., 5(2), pp. 20–25, 2024. IAEME
6. D. Chaudhary, S. Rajasegarar, and S. R. Pokhrel, "Towards Adapting Federated & Quantum Machine Learning for Network Intrusion Detection: A Survey," arXiv, 2025. arXiv
7. "Enhancing Privacy-Preserving Intrusion Detection through Federated Learning," Electronics, vol. 12, no. 16, p. 3382, 2023. MDPI
8. Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated Machine Learning: Concept and Applications," ACM Trans. Intell. Syst. Technol., vol. 10, no. 2, 2019. IAEME
9. R. C. Geyer, T. Klein, and M. Nabi, "Differentially Private Federated Learning: A Client-Level Perspective," Advances in Neural Information Processing Systems, vol. 30, 2017. IAEME
10. M. Abadi et al., "Deep Learning with Differential Privacy," in Proc. ACM SIGSAC Conf. Computer and Communications Security, 2016, pp. 308–318. IAEME
11. "Differential Privacy-Preserving Algorithms for Secure Training of Machine Learning Models," Int. J. Artif. Intell., Data Sci. & Machine Learn., 2025. IJADA Machine Learning
12. M. Sarhan, W. W. Lo, S. Layeghy, and M. Portmann, "HBFL: A Hierarchical Blockchain-based Federated Learning Framework for a Collaborative IoT Intrusion Detection," arXiv, 2022. arXiv
13. M. Rahmati, "Federated Learning-Driven Cybersecurity Framework for IoT Networks with Privacy-Preserving and Real-Time Threat Detection Capabilities," arXiv, 2025. arXiv
14. R. Sommer and V. Paxson, "Outside the Closed World: On Using Machine Learning for Network Intrusion Detection," in IEEE Symposium on Security and Privacy, 2010.
15. A. L. Buczak and E. Guven, "A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection," IEEE Commun. Surveys Tuts., vol. 18, no. 2, pp. 1153–1176, 2016.
16. C. Dwork, "Differential Privacy," in Automata, Languages and Programming, ICALP 2006, LNCS, vol. 4052, pp. 1–12.
17. H. B. McMahan et al., "Communication-Efficient Learning of Deep Networks from Decentralized Data," Proc. AISTATS, 2017.
18. P. Kairouz et al., "Advances and Open Problems in Federated Learning," Foundations and Trends® Mach. Learn., vol. 14, no. 1–2, pp. 1–210, 2021.
19. R. Shokri and V. Shmatikov, "Privacy-Preserving Deep Learning," in *Proc. ACM CCS*, 2015.
20. Y. Li, X. Huang, W. Yang, S. Wang, and Z. Zhang, "Federated Learning: Challenges, Methods, and Future Directions," IEEE Signal Process. Mag., vol. 37, no. 3, pp. 50–60, 2020.