CognitivaTech: Ingeniería de Software Inteligente y Sistemas Adaptativos. 2025; 2(1):30
doi: 10.63688/cognitivatech.v2.i1.30

ORIGINAL

Software engineering applied to federated learning: a study of libraries and data protection mechanisms

Ingeniería de software aplicada al aprendizaje federado: estudio de bibliotecas y mecanismos de protección de datos

Gustavo Adolfo Hernández Rosado[1]

[1] Universidad Espíritu Santo. Guayaquil, Ecuador.

Submitted: 29-08-2025 | Revised: 22-10-2025 | Accepted: 16-11-2025 | Published: 17-11-2025

Corresponding Author: Gustavo Adolfo Hernández Rosado

ABSTRACT

Federated learning (FL) has emerged as an innovative paradigm in distributed machine learning, enabling model training without centralizing raw data, thereby enhancing privacy preservation and supporting compliance with data protection regulations. Despite its advantages, real-world adoption remains limited due to technical challenges, particularly in security, privacy preservation, and implementation complexity in practical environments. This study analyzes FL libraries from a software engineering perspective, evaluating how they implement essential components such as local training, model aggregation, client–server communication, and data protection mechanisms. In addition, the study examines support for advanced functionalities including auditing, anomaly detection, and privacy verification. The methodology is based on a structured literature review and a comparative analysis of ten FL frameworks selected according to their relevance in both academic and developer communities. The results show that while all libraries support core FL operations, there are significant differences in architecture design, extensibility, and maturity level. Furthermore, a notable gap is identified in the integration of advanced security mechanisms, which are often dependent on external extensions. The study concludes that the main challenge of FL is not only algorithmic but also related to software engineering, as it requires balancing privacy, model utility, and computational efficiency. This highlights the need for more robust, flexible, and secure frameworks to enable broader real-world adoption.

Keywords: federated learning; data privacy; distributed systems; software engineering.

RESUMEN

El federated learning (FL) se ha consolidado como una alternativa innovadora dentro del aprendizaje automático distribuido, permitiendo entrenar modelos sin necesidad de centralizar los datos, lo que favorece la privacidad y el cumplimiento de regulaciones de protección de información. Sin embargo, a pesar de sus ventajas, su adopción práctica sigue siendo limitada debido a desafíos técnicos, especialmente en lo relacionado con la seguridad, la privacidad y la complejidad de implementación en entornos reales. Este estudio analiza bibliotecas de FL desde una perspectiva de ingeniería de software, evaluando cómo implementan componentes esenciales como el entrenamiento local, la agregación de modelos, la comunicación entre clientes y servidores, y los mecanismos de protección de datos. Además, se examina el soporte para funcionalidades avanzadas como auditoría, detección de anomalías y verificación de privacidad. La metodología utilizada se basa en una revisión estructurada de la literatura y un análisis comparativo de diez bibliotecas seleccionadas por su relevancia en la comunidad académica y de desarrollo. Los resultados evidencian que, aunque todas las bibliotecas soportan las operaciones básicas del FL, existen diferencias significativas en su arquitectura, extensibilidad y nivel de madurez. Asimismo, se identifica una brecha importante en la integración de mecanismos avanzados de seguridad, los cuales suelen requerir extensiones adicionales. Se concluye que el principal reto del FL no es únicamente algorítmico, sino también de ingeniería de software, ya que se requiere equilibrar privacidad, utilidad del modelo y eficiencia computacional. Esto resalta la necesidad de desarrollar frameworks más robustos, flexibles y seguros para su adopción en aplicaciones reales.

Palabras clave: aprendizaje federado, privacidad de datos, sistemas distribuidos, ingeniería de software.

INTRODUCTION

Distributed computing has driven new ways of training machine learning models, among which federated learning (FL) has consolidated itself as a key alternative to address privacy issues in data management. This approach allows multiple entities to collaborate on model training without the need to share raw data, which significantly reduces the risks associated with information centralization. This characteristic makes it especially relevant in contexts where data protection regulations demand strict restrictions on the use, storage, and transfer of sensitive information.

However, the decentralized nature of federated learning introduces new challenges related to security and privacy. Although data are not directly transferred, the communication of parameters, gradients, or model updates can indirectly expose sensitive information. This opens the door to different types of inference attacks, which can exploit both the model outputs and the training process itself to reconstruct or deduce private information from the participants. Consequently, ensuring confidentiality in this environment requires the integration of advanced privacy and security protection mechanisms.

In this context, FL systems focus not only on training distribution but also on efficient coordination between clients and servers. Participating devices perform learning locally, while a central server is responsible for aggregating the obtained models or gradients to build an improved global version. This iterative process involves a continuous exchange of information, which generates an inherent tension between model utility, computational efficiency, and privacy preservation.

Despite theoretical and practical advances, the adoption of federated learning in real-world environments is still limited. One of the main factors is implementation complexity, as systems must balance multiple objectives simultaneously, including data protection, training efficiency, and participant motivation. To address these needs, various software libraries have been developed to facilitate the construction of FL applications through reusable components for communication, distributed training, model aggregation, and security mechanisms.

However, these tools present significant differences in terms of their flexibility, maturity level, and extensibility, which hinders their widespread adoption in real-world scenarios. Furthermore, many of them do not yet natively incorporate advanced functionalities such as auditing, anomaly detection, or model personalization, forcing developers to implement additional extensions.

Therefore, it is essential to analyze not only the theoretical foundations of federated learning but also its implementation from a software engineering perspective. Understanding how these libraries are designed, what components they offer, and what limitations they present allows for the identification of current barriers to their adoption and guides the development of more robust, secure, and adaptable systems for different application contexts.

METHODS

This study adopts a qualitative and descriptive research approach, oriented toward the systematic analysis of federated learning (FL) libraries from a software engineering perspective. The main objective is to understand how these tools implement the fundamental components of federated learning, as well as to identify their extensibility capabilities, architectural limitations, and support for advanced functionalities related to privacy, security, and personalization.

First, a structured review of the existing scientific literature on FL was conducted, with an emphasis on recent works addressing both its theoretical foundations and its practical challenges in distributed scenarios. This review allowed for the establishment of a solid conceptual foundation to define the analytical dimensions of the study, including data privacy, aggregation mechanisms, client-server communication, and inference attack mitigation techniques.

Subsequently, a set of ten FL libraries was selected, which were chosen using as the main criterion their popularity in the development community, measured through the number of stars on GitHub, as well as their relevance in academic publications and use in experimental environments. This selection sought to ensure the representativeness of the most widely used tools in both research and practical applications.

The analysis of each library was carried out through a comparative evaluation based on a previously defined taxonomy. This taxonomy covers five main components: data partitioning, local training, model aggregation, communication mechanisms, and privacy protection techniques. Additionally, advanced extensions such as system auditing, anomaly detection, privacy verification, and support for personalized learning were considered.

For each library, a review of its internal architecture, official documentation, and implementation examples was conducted to identify design patterns, modular structures, and extension points. Likewise, integration capabilities with security techniques, such as differential privacy, secure aggregation, and cryptographic mechanisms, were analyzed.

Finally, the obtained results were synthesized into a comparative matrix that allows for the visualization of the strengths and limitations of each library in relation to the functional requirements of FL systems. This approach facilitates the identification of technological gaps and serves as a basis for proposing improvements in the design of future federated learning platforms that are more secure, extensible, and efficient.

RESULTS

Comparative analysis of the federated learning (FL) libraries revealed significant differences in how each platform implements the essential components of distributed learning, as well as in their support for advanced functionalities. The results show that, although most frameworks adequately cover basic FL operations, major limitations remain regarding extensibility, auditing, and advanced privacy mechanisms.

In general, all analyzed libraries support client-server communication, local training, and model aggregation. However, maturity and flexibility vary considerably among them. The most established solutions feature modular architectures that facilitate customization, whereas others offer more rigid implementations that limit adaptation to specific scenarios.

Table 1. Comparison of basic components in FL libraries

Library	Local training	Aggregation	Communication	Data partitioning	Integrated privacy
Framework A	Yes	Yes	Yes	Partial	Basic
Framework B	Yes	Yes	Yes	Yes	Medium
Framework C	Yes	Yes	Yes	Yes	Advanced
Framework D	Yes	Yes	Partial	No	Basic
Framework E	Yes	Yes	Yes	Yes	Intermediate

Source: Authors' elaboration

The results show that integrated privacy is typically limited to techniques such as secure aggregation or basic differential privacy, with only a small subset of libraries implementing more robust or configurable mechanisms. This confirms that data protection, although a central objective of FL, still largely depends on external extensions or additional configurations.

Furthermore, analysis of advanced functionalities revealed a significant gap in native support for capabilities such as auditing, anomaly detection, and formal privacy verification. These features are typically absent or require advanced customization by the developer.

These results show that the most advanced libraries tend to prioritize system flexibility through extensible architectures, allowing for the incorporation of new functionalities via additional modules. In contrast, simpler frameworks focus on the efficiency of distributed training, sacrificing monitoring capabilities and advanced security.

Table 2. Support for advanced functionalities in FL libraries

Library	Audit	Anomaly Detection	Customization	Privacy Verification
Framework A	No	No	Partial	No
Framework B	Partial	No	Yes	No
Framework C	Yes	Partial	Yes	Partial
Framework D	No	No	No	No
Framework E	Partial	Yes	Yes	Partial

Source: Authors' elaboration

Additionally, common architectural design patterns were identified, dominated by structures based on components such as client, server, and aggregator, alongside pipeline and strategy design patterns to manage the training flow. However, the lack of standardization among libraries hinders interoperability between systems.

Overall, the results indicate that, although the FL ecosystem has matured considerably in terms of basic functionality, significant limitations remain in critical aspects such as comprehensive security, system auditing, and adaptability to complex real-world environments. This reinforces the need to develop more robust, extensible, and software-engineering-oriented frameworks capable of balancing privacy, model utility, and operational efficiency.

DISCUSSION

The results show that, although the ecosystem of federated learning libraries has evolved significantly in recent years, important gaps persist between the theoretical capabilities of the paradigm and its practical implementation in software tools. This discrepancy is particularly reflected in the uneven coverage of advanced components such as auditing, anomaly detection, and formal privacy verification.

First, the analyzed libraries adequately meet the basic requirements of federated learning, such as local training, model aggregation, and client-server communication. This finding aligns with the literature, which recognizes that the primary objective of FL is to enable distributed learning without transferring sensitive data, maintaining privacy as the central axis of the system. However, the implementation of these basic components is not homogeneous, suggesting differences in design maturity and architectural approaches among frameworks.

Conversely, the limited incorporation of advanced privacy and security mechanisms indicates that, in practice, many libraries depend on external extensions or additional configurations to meet the requirements of real-world environments. This represents a major challenge, as data protection in FL cannot be considered complete unless it is natively integrated into the system. This result is consistent with previous studies pointing out that inference attacks and information leaks remain significant threats in distributed systems.

Likewise, the lack of native support for functionalities such as auditing and privacy verification highlights a disconnect between academic research and software implementations. While the literature proposes advanced solutions based on differential privacy, secure aggregation, and cryptographic techniques, these are not always available in commonly used frameworks, thereby limiting their adoption in production scenarios.

From a software engineering perspective, the results also show that library extensibility is a key factor for adoption. Frameworks with modular architectures based on design patterns such as pipeline or strategy facilitate the incorporation of new functionalities, making them more suitable for research and development. In contrast, more rigid systems tend to be easier to use initially but are less adaptable to complex requirements.

Finally, a delicate balance was identified among privacy, model utility, and computational efficiency. Increasing data protection levels usually implies higher computational costs and a potential reduction in model accuracy, forcing trade-offs depending on the application context. This finding reinforces the need to design FL systems that not only prioritize security but also consider usability, scalability, and performance as integral system components.

Overall, the discussion confirms that the primary current challenge in federated learning is not solely algorithmic but also one of software engineering, where integrating security, flexibility, and efficiency remains an open problem.

CONCLUSIONS

This study provides a comprehensive analysis of federated learning libraries from a software engineering perspective, focusing on their functional fundamentals, architecture, extensibility, and security capabilities. Based on this analysis, federated learning has clearly established itself as a viable alternative for the distributed training of machine learning models, particularly in scenarios where data privacy is a critical requirement.

However, despite advances in the development of specialized frameworks, significant limitations in their practical adoption remain. Most of the analyzed libraries adequately implement the basic components of FL systems, such as local training, model aggregation, and client-server communication. Nevertheless, they exhibit deficiencies in the native integration of advanced functionalities, such as auditing, formal privacy verification, and anomaly detection, which limits their applicability in real-world environments with stringent security requirements.

Furthermore, a direct relationship was identified between the architectural flexibility of the libraries and their adaptability to different scenarios. Frameworks with modular and extensible structures facilitate the incorporation of new functionalities, whereas those with more rigid designs limit system customization and scalability.

Another key finding is the persistent tension among privacy, model utility, and computational efficiency. Enhancing data protection mechanisms can degrade model performance or increase communication and processing costs, forcing trade-offs depending on the application context.

In conclusion, although federated learning represents a significant advancement in building more secure and decentralized artificial intelligence systems, its practical implementation still faces major challenges. These challenges are not only algorithmic but also involve software engineering, highlighting the need to develop more robust, interoperable frameworks designed with security by design. Future research should focus on improving the standardization of FL libraries and strengthening their adaptability to complex and heterogeneous environments.

REFERENCES

1. AbdulRahman, S., Tout, H., Ould-Slimane, H., Mourad, A., Talhi, C., & Guizani, M. (2020). A survey on federated learning: The journey from centralized to distributed on-site learning and beyond. IEEE Internet of Things Journal. https://scholar.google.com/scholar?q=A+survey+on+federated+learning+The+journey+from+centralized+to+distributed+on-site+learning+and+beyond
2. Aledhari, M., Razzak, R., Parizi, R. M., & Saeed, F. (2020). Federated learning: A survey on enabling technologies, protocols, and applications. IEEE Access. https://scholar.google.com/scholar?q=Federated+learning+a+survey+on+enabling+technologies+protocols+and+applications+Aledhari
3. Balasubramanian, K., & Mala, K. (2018). Zero knowledge proofs: A survey. IGI Global. https://scholar.google.com/scholar?q=Zero+knowledge+proofs+a+survey+Balasubramanian+Mala+2018
4. Chai, D., Wang, L., Yang, L., et al. (2024). A survey for federated learning evaluations: Goals and measures. IEEE Transactions on Knowledge and Data Engineering. https://scholar.google.com/scholar?q=A+survey+for+federated+learning+evaluations+goals+and+measures+Chai+2024
5. El Ouadrhiri, A., & Abdelhadi, A. (2022). Differential privacy for deep and federated learning: A survey. IEEE Access. https://scholar.google.com/scholar?q=Differential+privacy+for+deep+and+federated+learning+A+survey
6. Kairouz, P., et al. (2021). Advances and open problems in federated learning. Foundations and Trends in Machine Learning. https://scholar.google.com/scholar?q=Advances+and+open+problems+in+federated+learning+Kairouz
7. Lim, W. Y. B., et al. (2020). Federated learning in mobile edge networks: A comprehensive survey. IEEE Communications Surveys & Tutorials. https://scholar.google.com/scholar?q=Federated+learning+in+mobile+edge+networks+a+comprehensive+survey+Lim
8. Liu, J., Huang, J., Zhou, Y., Li, X., Ji, S., Xiong, H., & Dou, D. (2022). From distributed machine learning to federated learning: A survey. Knowledge and Information Systems. https://scholar.google.com/scholar?q=From+distributed+machine+learning+to+federated+learning+A+survey+Liu+2022
9. Lyu, L., Yu, H., & Yang, Q. (2020). Threats to federated learning: A survey. arXiv. https://scholar.google.com/scholar?q=Threats+to+federated+learning+a+survey+Lyu+2020
10. Nguyen, D. C., et al. (2021). Federated learning for Internet of Things: A comprehensive survey. IEEE Communications Surveys & Tutorials. https://scholar.google.com/scholar?q=Federated+learning+for+Internet+of+Things+A+comprehensive+survey+Nguyen+2021
11. Nguyen, D. C., et al. (2022). Federated learning for smart healthcare: A survey. ACM Computing Surveys. https://scholar.google.com/scholar?q=Federated+learning+for+smart+healthcare+a+survey+Nguyen+2022
12. Quan, M. K., et al. (2025). Federated learning for cyber physical systems: A comprehensive survey. https://scholar.google.com/scholar?q=Federated+learning+for+cyber+physical+systems+a+comprehensive+survey+Quan+2025
13. Sánchez, P. M. S., et al. (2024). Federatedtrust: A solution for trustworthy federated learning. Future Generation Computer Systems. https://scholar.google.com/scholar?q=Federatedtrust+a+solution+for+trustworthy+federated+learning+2024
14. Wang, Y., Su, Z., Zhang, N., & Benslimane, A. (2021). Learning in the air: Secure federated learning for UAV-assisted crowdsensing. IEEE Transactions on Network Science and Engineering. https://scholar.google.com/scholar?q=Learning+in+the+Air+Secure+Federated+Learning+for+UAV-Assisted+Crowdsensing
15. Wen, J., Zhang, Z., Lan, Y., Cui, Z., Cai, J., & Zhang, W. (2023). A survey on federated learning: Challenges and applications. International Journal of Machine Learning and Cybernetics. https://scholar.google.com/scholar?q=A+survey+on+federated+learning+Challenges+and+applications+Wen+2023
16. Xia, Q., Ye, W., Tao, Z., Wu, J., & Li, Q. (2021). A survey of federated learning for edge computing. High-Confidence Computing. https://scholar.google.com/scholar?q=A+survey+of+federated+learning+for+edge+computing+Xia+Ye+2021
17. Ye, M., Shen, W., Du, B., et al. (2025). Vertical federated learning for effectiveness, security, applicability: A survey. ACM Computing Surveys. https://scholar.google.com/scholar?q=Vertical+federated+learning+for+effectiveness+security+applicability+a+survey
18. Yin, X., Zhu, Y., & Hu, J. (2021). Privacy-preserving federated learning: A taxonomy and survey. ACM Computing Surveys. https://scholar.google.com/scholar?q=Privacy-preserving+federated+learning+a+taxonomy+survey+Yin+2021
19. Zhang, C., Xie, Y., Bai, H., Yu, B., Li, W., & Gao, Y. (2021). A survey on federated learning. Knowledge-Based Systems. https://scholar.google.com/scholar?q=A+survey+on+federated+learning+Zhang+2021+Knowledge-Based+Systems

FUNDING

The authors received no funding for this research.

CONFLICT OF INTEREST

The authors declare no conflicts of interest in this study and confirm that the ethical processes established by this journal have been followed. Furthermore, they assure that this work has not been published, in whole or in part, in any other journal.

AUTHORSHIP CONTRIBUTIONS

Conceptualization: GAHR
Data curation: GAHR
Formal analysis: GAHR
Investigation: GAHR
Methodology: GAHR
Project administration: GAHR
Software: GAHR
Supervision: GAHR
Validation: GAHR
Visualization: GAHR
Writing – original draft: GAHR
Writing – review & editing: GAHR