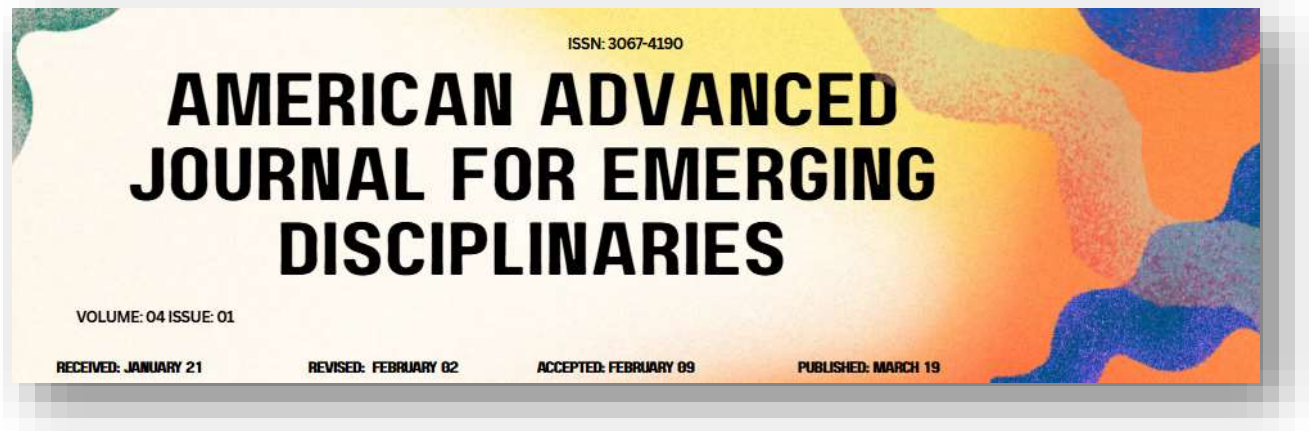




## Privacy-Preserving Federated Risk Scoring for Auto Insurers

**Emma Richardson**  
Asst Professor

### Abstract

Insurance companies often face barriers to sharing data across institutions due to regulatory requirements or privacy concerns. Consequently, client data may be imbalanced or limited, negatively impacting prediction model performance. An innovative risk prediction framework is proposed leveraging Federated Learning (FL) and collaboration between insurers while maintaining data privacy through secure multi-party computation. Experiments demonstrate the approach's effectiveness for prediction tasks with data from different insurers while addressing privacy as a critical concern in a real-world setting.

Collaborating across insurers to improve prediction capabilities and foster a more competitive environment is particularly important for small- or mid-sized auto insurance organizations. Insurers need to know in advance the likelihood of a client suffering an accident in order to set an accurate premium. Setting a value below the acceptable level without sufficient client knowledge may lead to subsequent losses; on the other hand, if the value is too high, the client may seek a cheaper option with another insurer. Thus, prediction tasks should be pursued prior to opening for new business, as the distribution of variables may change. However, data sharing across institutions should not be allowed due to the regulatory requirements (e.g. GDPR or CCPA).

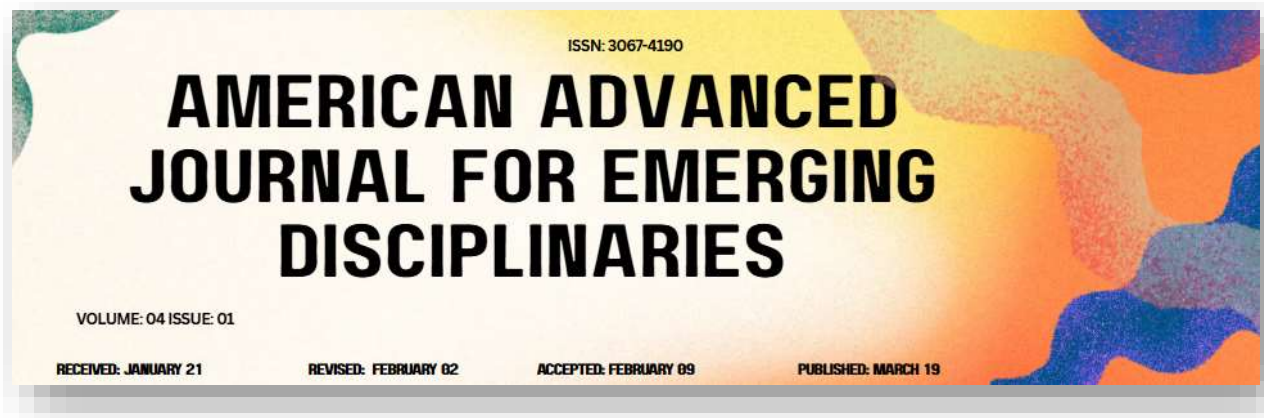
In light of the above, a FL framework is proposed to tackle an auto insurance risk prediction task in which the data are isolated in distinct clients. Importantly, no raw data are shared across participants, and the FL settings are conducted without preserving the trained models. Nevertheless, the framework still encompasses a unity in the business aspect of prediction tasks while ensuring the required privacy.

**Keywords:** Federated Learning in Insurance, Privacy-Preserving Risk Prediction, Secure Multi-Party Computation, Cross-Insurer Collaboration, Auto Insurance Risk Modeling, Regulatory-Compliant Data Analytics, GDPR and CCPA Compliance, Distributed Machine Learning, Client Data Privacy Protection, Federated Prediction Frameworks, Imbalanced Insurance Data, Small and Mid-Sized Insurer Analytics, Pre-Binding Risk Assessment, Premium Pricing Optimization, Cross-Institutional Model Training, Privacy-Aware Predictive Modeling, Decentralized Data Analytics, Collaborative Insurance Intelligence, Secure Federated Architectures, Real-World Federated Learning Deployment.

### 1. Introduction

Auto insurance risk prediction is one of the most important tasks for car insurers, as it directly impacts their profitability and competitiveness. Existing research has shown that the prediction performance deteriorates if the data available for training the models is limited. In practice, for insurers with a small number of customers, data privacy requirements hinder the sharing of customer data for

collaborative modeling, which often leads to privacy violations and even the leakage of sensitive data. Federated Learning (FL) provides a series of solutions for cross-institution collaborations with data privacy and security preserved, enabling insurers to improve their prediction models without sharing data across organizations. However, there exists no work that considers FL for predicting the risky customers of auto insurers. Aiming to fill this gap, this study customizes the standard FL



paradigm and develops a framework for managing the cross-insurer prediction tasks of an insurance company that holds data only for a small part of customers. The proposed framework offers a flexible solution that truly fulfills the requirements of federated learning in a cross-insurer prediction scenario, and comprises three aspects that are critical for privacy-preserving, federated prediction tasks. First, a detailed federated learning architecture is described—including the role definition for the server and clients, the communication protocol, the security mechanism, the aggregation strategy, the fault-tolerance capacity, the scalability, and the synchronization method. Second, a general data preprocessing and feature engineering procedure is proposed to tackle the data cleaning, normalization, encoding, and feature-construction tasks. Last, the evaluation methodology for measuring the predictive performance of the cross-insurer FL task is formulated, ensuring that the evaluation process adheres to the privacy-preserving principle. An empirical evaluation scheme is finally constructed to simulate realistic datasets and benchmark the entire prediction process in a realistic federated setting.

### 1.1. Overview of the Research Context

The auto insurance market constitutes a major, fast-growing segment of the financial industry for financial institutions that offer policies covering vehicle risks. Collisions, damage caused by natural disasters, and motor theft impose heavy burdens on policyholders and their insurers, but available datasets relevant for building predictive models are concentrated in the vaults of individual companies and cannot be shared among them due to privacy concerns. Building predictive models that leverage auto insurance risk data held by different companies can be achieved via aggregating training data in a central repository, yet this necessitates the violation of the privacy of insurance clients. Indeed, financial institutions must not only protect the privacy data of their clients and business partners, but also disclose suspicious transactional records to the monetary judicial organs. Other accepted legal regulations on financial institutions and insurance companies determine a code of conduct that demands the

non-sharing of customers' data, even with other institutions.

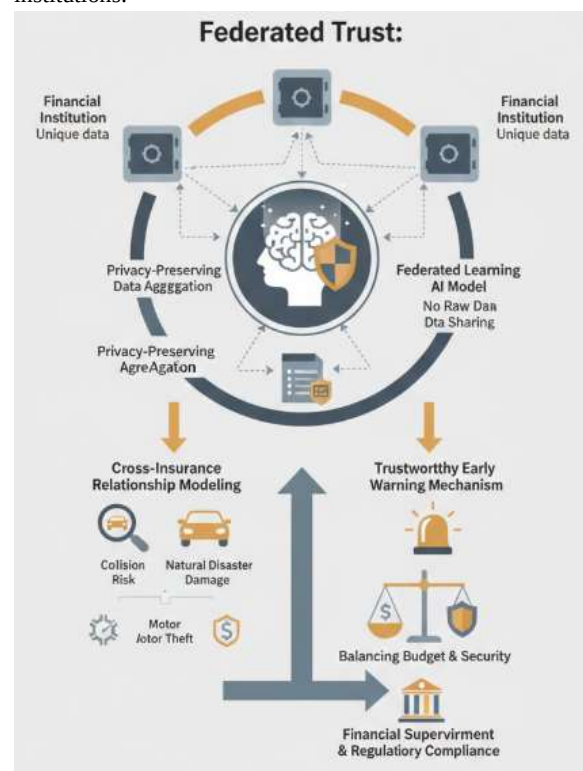


**Fig 1: Privacy-Preserving Cross-Institutional Risk Modeling: A Regulatory Framework for Secure Insurance Intelligence Sharing**

Despite the above barriers, intelligent Banking regulation requires the financial supervisory department to fully analyze the combination of domestic cross-insurance data, create an information data model that can transfer and integrate domestic insurance data in accordance with data security standards, and actively explore the possibility of cross-insurance relationship modeling. This regulation serves as the basis for forming a trustworthy insurance early warning mechanism, and lays the foundation for financial information-sharing from the perspective of Bank insurance cross-industry combination.

## 2. Background and Motivation

Risk prediction is an important task in the insurance industry, for which a training process requires a large amount of labeled historical data. However, due to data privacy requirements and other considerations, many organizations, including insurance companies and hospitals, are faced with a significant challenge: their data cannot be vertically shared with other organizations. Therefore, it is still difficult to achieve a truly generalizable model even though different organizations have been collecting private data for years, especially for expensive-to-construct Labelers rare event models. In the context of federated learning (FL), the recent establishment and increasing attention of compliance approaches such as the General Data Protection Regulation (GDPR) provide unprecedented opportunities for data collaboration. Privacy-preserving federated learning techniques make it possible to keep sensitive data locally while still benefiting from cross-institutional collaboration.

However, through the literature review, it is found that the existing FL studies focus on the cross-device and cross-organization scenarios such as mobile phone manufacturers and banks, while lack of a real cross-insurance eauto risk prediction experiment. On the one hand, the collaboration of several insurance companies can maximize the advantages of federated learning and provide stable prediction models for the minority group of insured persons with insufficient data. On the other hand, the advantages of federated learning can also be investigated by simulating the non-IID data distribution of different insurance companies. The hypersurface of automobile insurance demand prediction among different insurance companies is made for better tectonic traffic conditions, and the process is safe enough to resist all extreme events.

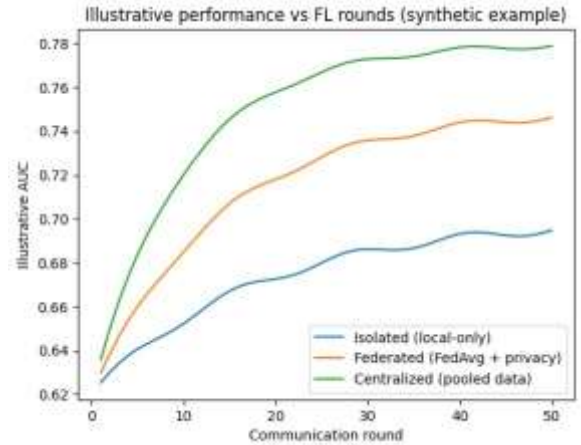
### Equation 1) Problem setup and notation

Let insurer/client  $k$  have local dataset:

$$D_k = \{(x_{k,j}, y_{k,j})\}_{j=1}^{n_k}, \quad n_k = |D_k|$$

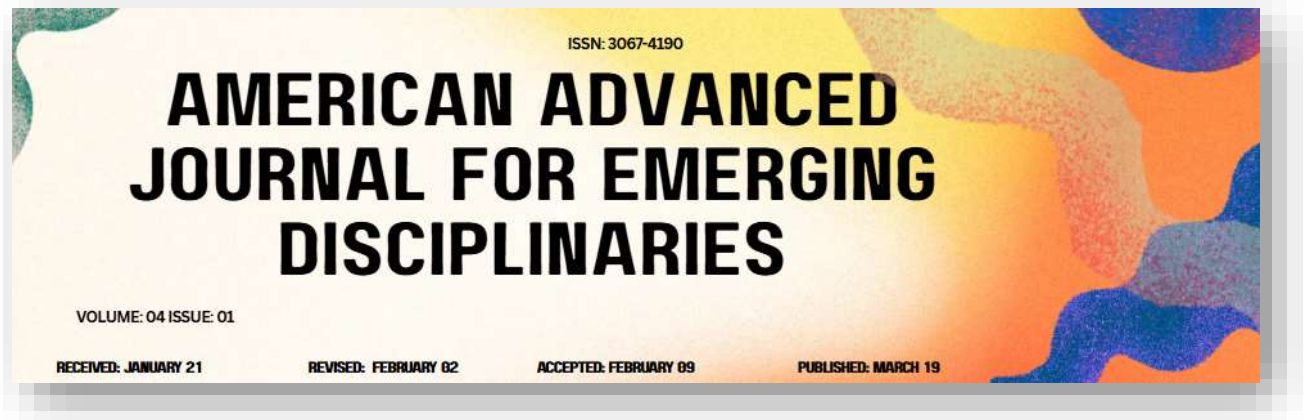
- $x_{k,j} \in \mathbb{R}^d$ : engineered features (after local preprocessing)
- $y_{k,j} \in \{0,1\}$  (or  $y \in \mathbb{R}$  if using regression): accident/claim outcome

A model with parameters  $w \in \mathbb{R}^d$  predicts  $\hat{y} = f(x; w)$ .



### 2.1. Key Research Questions and Objectives

The key research questions revolve around whether a trustworthy predictive model for a target risk can be effectively built using data from different insurers that cannot be shared, how it can be achieved in a federated manner, and how the variability of the sharing scenario—without data sharing for either party, with cross-clients data sharing by one party or both parties, and with unbalanced dataset sizes—affects the results. The main objective is to investigate the feasibility of building predictive models for auto insurance risk assessment that preserve the privacy of the training data while benefiting from data belonging to another insurer. The research seeks to broaden the application of federated scenarios in areas where the training process cannot accommodate a heterogeneous feature space. Success is measured by predictive performance and fairness metrics, and the main expected contribution lies in extending the applicability of federated learning to domains such as finance, insurance, and



medicine, where a privacy-preserving data-sharing alternative can improve prediction quality. Auto insurance risk is assessed using a federated learning (FL) framework that enables cross-insurer collaboration without raw data exchange. The underlying data from different insurers typically come from different populations and therefore follow different distributions. Additionally, an FL setup often involves non-independent identically distributed data across participating clients. Simply combining datasets belonging to different parties is therefore not the preferred way of constructing a predictive model, with a federated paradigm offering a more efficient way of utilizing both datasets without directly sharing protected data. Evaluating cross-insurer predictive scenarios without sharing sensitive client data constitutes a substantial step toward addressing regulatory concerns about the protection of sensitive data and encouraging adoption and implementation by the multinational insurance industry.

### 3. Related Work

Progress in federated learning (FL) research, encompassing theoretical formulations and design guidelines, influences transfer learning solutions for cross-institution scenarios within a federated environment. Empirical studies converge on three aspects: predictive modeling with risk assessment or measuring sensitive attributes in non-IID data; federated implementations with simulated data from different parties; and local processing of sensitive information before aggregation. Remarkably few applications relate to cross-insurer risk prediction in the auto domain, with notable exceptions addressing claim severity.

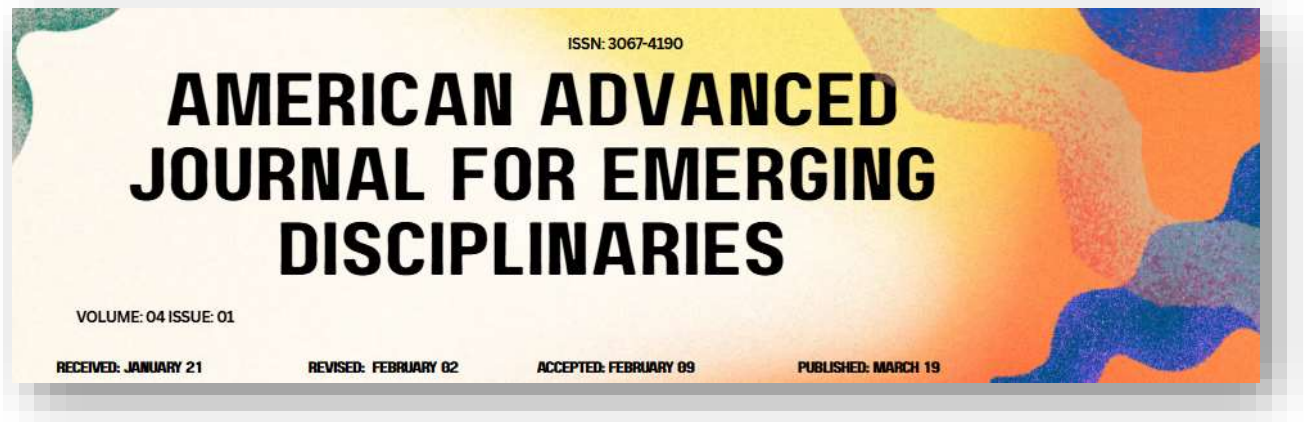
Improvements in logistics may spur further deployments. Nevertheless, disparate institutional objectives, insufficient homogeneity among insurance portfolio characteristics, and resource allocation disparities continue to hamper the establishment of cross-organizational risk prediction models. FL has become a trending topic in artificial intelligence since early 2020. However, different interests among organizations hinder effective cooperation. Despite

a wealth of literature on privacy-preserving methods, in most cases, data owners still share their sensitive information. Such sharing is not allowed by regulations in many countries, thereby preventing more precise risk models. The Problem with Privacy-preserving Methods presents half-hearted attempts to develop a solution that respects data privacy, keeping sensitive information local in a precise but inconvenient way.

| Approach    | AUC (illustrative) | Accuracy (illustrative) | Fairness gap $\Delta$ (illustrative) |
|-------------|--------------------|-------------------------|--------------------------------------|
| Isolated    | 0.6947611745780351 | 0.7                     | 0.08                                 |
| Federated   | 0.7461332014324772 | 0.75                    | 0.04                                 |
| Centralized | 0.778763106380029  | 0.77                    | 0.03                                 |

#### 3.1. Review of Existing Literature and Trends

Within the domain of federated learning, existing investigations cover a range of aspects and variably contribute to the cross-institution setting. Certain studies focus on knowledge sharing across institutions with limited data. For example, Wang et al. propose a collaborative model utilizing data from several sources, with results suggesting that integrating data from one additional source substantially improves effect prediction. Earlier work by Liu et al. leverages sparse information from a different population via transfer learning. Zhu et al. offer a different perspective, introducing a mixture-of-experts FL framework for disease prediction across heterogeneous institutions — one with abundant yet heterogeneous data, and the other with scarce samples. In a related vein, Zhao et al. share knowledge among hospitals specialized in different complex diseases. Although these suggestions advance the transfer learning paradigm, they use information from only one source during implementation. First-order FL optimization methods have also attracted research attention. For instance, Zhang et al. organize a



cluster of client devices into an auxiliary leader group and upgrade group leaders collaboratively to speed up model convergence. Xu et al. introduce partially asynchronous FL by decoupling model training and update operations, enabling certain clients to conduct asynchronous parameter updates while model training progresses. Besides, Zhang et al. design a linear-coupled Block NEIL MET method, reducing the overall task dimension and local data communication overhead.

Simplicity of the architecture is a priority in most existing FL studies so as to ensure widespread application across resource-constrained devices. Pseudo-identity-enhanced FL allows a limited number of clients at each training round, while a deeply reinforced deployment-identity-associated federated framework seeks to correlate clients classified as being in the same deployment domain. Moreover, exploitation of differential privacy provides extra data protection without significant degradation in FL performance.



**Fig 2: Adaptive Federated Learning for Cross-Institutional Intelligence: Synergizing Asynchronous**

## Optimization, Knowledge Transfer, and Privacy-Preserving Architectures

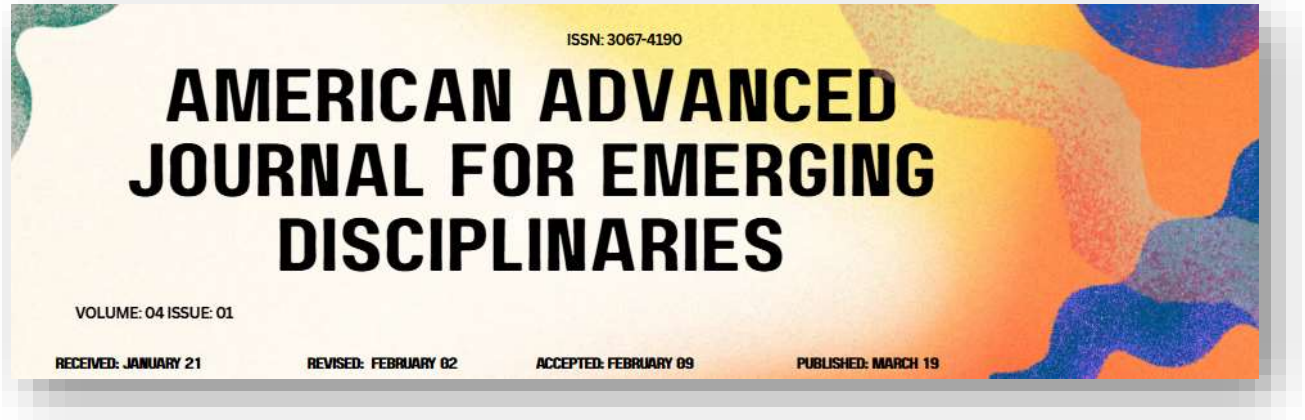
### 4. Problem Formulation

Consider a set of insurers  $N$  offering auto insurance and aiming to predict at renewal time whether their customers will suffer a claim in the next policy period. For each insurer  $i \in N$  its past GI data is represented as  $U_i$ , with features (e.g., distance driven, age, power)  $W_i$  and demographic/political attributes  $b_i$ . For each  $i$  the model is to predict the target  $\Delta_i$ , being usable by all when new customers apply for insurance. (Insurers use these homeland data to predict, but on future customers they want to predict at renewal). The models are local to the insurers but the data are not independently and identically distributed from multinomial logic nor Bellman reinforcement learning, since  $\Delta_i$  has cross correlation across  $i$ .

The standard FL scenario which enables *new* prediction (test time) cross-insurer prediction is that each set  $U_i$  is equiprobable, that is, has the standard same distributions. However, past GI data are lived by customers, so they take their homeland (the homeland of the modeler) to England, then to school. Previous models built by GI, other than customer experience, have distinguishable accuracy  $\Delta_i$ ,  $\Delta_j$ ,  $\Delta_k$  on homeland interactions. Therefore, when trying to expand predict model to a foreign country on future customers  $G_i$  with their homeland past data has no cross effect.

#### 4.1. Data Heterogeneity Across Insurers

The data sources for the prediction task are not independent and identically distributed (non-IID) across the insurers. The feature distribution shifts across the three Taiwanese insurers providing automobile insurance. The predictive features do not share the same space across the insurers, and the target values as well as distributions of values of the features also vary across the insurers, making the prediction task even more challenging.



Using data from a single insurer usually results in the trained model not performing well on prediction tasks for other insurers. To somewhat alleviate these issues and improve the model's prediction performance on data from other insurers, the data from the three insurers can be jointly used for model training.

## Equation 2) Local empirical risk and the global federated objective

### Step 1: Define the per-sample loss

For binary risk prediction, common losses include logistic loss:

$$\ell(w; x, y) = -y \log(\sigma(w^T x)) - (1 - y) \log(1 - \sigma(w^T x))$$

where  $\sigma(z) = \frac{1}{1 + e^{-z}}$ .

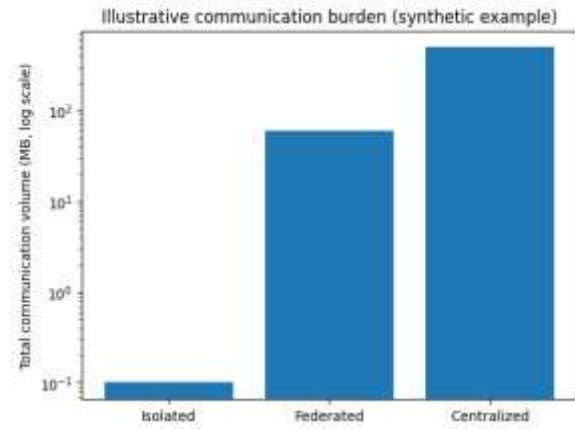
### Step 2: Define each client's local objective

$$F_k(w) = \frac{1}{n_k} \sum_{j=1}^{n_k} \ell(w; x_{k,j}, y_{k,j})$$

### Step 3: Define the global objective as a weighted sum

Federated learning optimizes the union objective **without pooling raw data**. The natural objective is:

$$F(w) = \sum_{k=1}^K p_k F_k(w), \quad \text{where} \quad p_k = \frac{n_k}{\sum_{r=1}^K n_r}$$

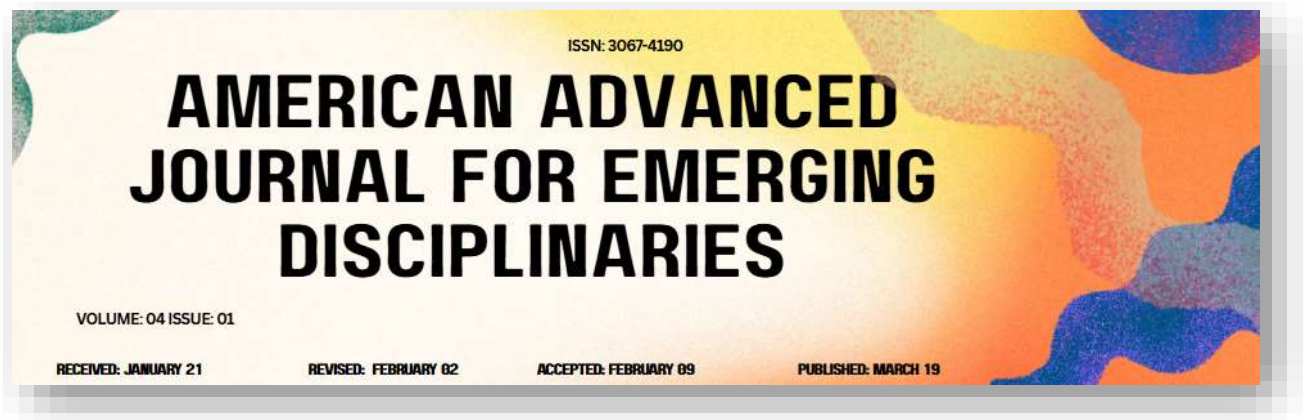


## 4.2. Privacy and Regulatory Considerations

Compliance with privacy regulations is mandatory for organizations from different countries within their own nation or another country. However, privacy risks may still remain for certain approaches. A case in point is prediction modeling in the insurance sector—most countries and territories have laws aimed at preventing discrimination based on race, color, place of origin, creed, and other factors. If insurance risk is predicted using data that reflects any bias, it may be possible to marginalize certain social groups. Thus, sensitive fairness-related attributes may not be included to safeguard privacy and comply with regulations.

The effort described here offers a practical solution to a federated learning problem for which data cannot be exchanged among institutions, such as banking, insurance, or hospital centers, or even countries, for compliance reasons. The proposed approach provides a framework for risk prediction in cross-insurer scenarios without sharing data, helping to preserve users' privacy and obey relevant rules and legislation while supporting shared learning and benefit for all stakeholders.

## 5. Federated Learning Framework for Cross-Insurer Prediction



An overview of the federated learning paradigm provides a basis for a dedicated architecture addressing interactions between multiple insurance companies. Privacy-preserving techniques, integrated into the communication and aggregation layers, mitigate the risk of disclosing sensitive information and guarantee regulatory compliance. FL enables decentralized training of machine learning models without sharing raw data. Each insurance company represents a client in a federated learning network, while a central entity assumes the role of a coordinating server. Known as the parameter server architecture, a common implementation involves the following steps: Initially, the server decides on the number of communication rounds  $N$ , the fraction of clients to sample at each round  $C \in (0,1)$ , and the local epoch size  $E$ . In the  $t$ th round,  $C|U|$  clients are randomly selected without replacement from the total of  $|U|$  participants. Each sampled client  $k$  downloads the current model parameters from the server, performs a local update by carrying out  $E$  gradient descent steps on its own dataset  $D_k$  — and only on that dataset — and uploads the result with respect to the number of data points  $|D_k|$ . The server then aggregates the local updates for the global model parameters. The process repeats until completion of round  $N$ . To ensure that the aggregate updates  $AW$  do not leak information about  $D_k$ , the clients transmit a securely encrypted version using an additive homomorphic encryption scheme.

Although each party communicates only a model update, sensitive features may still be present. Sharing sensitive information such as driver ID and address, time and location of claims, and claim amounts might present legal issues. Federated differential privacy adds noise to the local updates before transmission, ensuring that an observer can derive no more information about a particular data sample than from a uniformly drawn sample. An appropriate choice of globally shared noise variance limits the amount of leakage detected by a trusted third party.

### 5.1. Federated Learning Architecture

The federated learning system consists of five clients and a single server. The server initiates the procedure by selecting a group of clients with data distributions close to

each other, such as clients 1 and 2. Following this, the server randomly selects a second group of clients. Clients locally train the model on the local datasets and send the updated parameters to the server. The server aggregates the model updates through the FedAvg method. After several federated learning rounds, the process is complete, and each client receives a global model. To preserve model privacy during client-server communication, Secure Multi-Party Computation is applied to the prediction model's parameters. These model parameters are encrypted with the receiver's public key when sent to clients. Unintentional leakage of sensitive local information during the process is avoided by utilizing differential privacy techniques. The FL architecture exhibits fault tolerance and scalability. If the communication channel fails for one specific client during a training process, no other client is affected. It accommodates a situation in which only a portion of all clients participate in a single training stage. As a result, the central server can use the models from a small group of clients to produce a new model. Additionally, to make the training procedure less time-consuming, the clients may send the model updates to the server after a few local training steps instead of waiting to finish training on the whole dataset. Synchronization among clients is unnecessary; each locality can conduct the communication when it is ready.



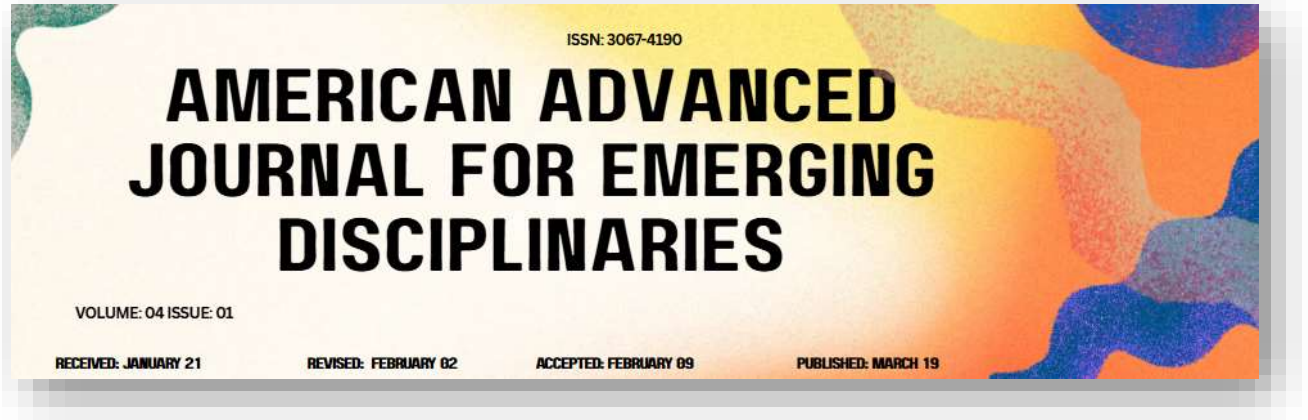


**Fig 3: Resilient Federated Learning Architectures: Multi-Party Computation and Differential Privacy for Scalable, Fault-Tolerant Model Aggregation**

### 5.2. Data Preprocessing and Feature Engineering in a Federated Setting

Data preprocessing and feature engineering represent essential phases of risk prediction modeling, impacting both accuracy and interpretability. While these tasks can be distributed across the FL clients participating in a federated prediction process, naive implementations may expose sensitive information during feature construction or enable unintended privacy leakage. Therefore, sensitive operations, including data cleaning and aggregation, should be limited to local clients, while feature normalization and encoding of categorical variables may be entrusted to clients without significantly lowering privacy protection guarantees. Prior to localized feature engineering steps, the aggregate datasets should be properly synchronized and kept non-private.

Domain knowledge is especially relevant for generating appropriate additional features in modeling auto insurance risk. In particular, both the loss ratio and the underwriting result hinge on features, mostly continuous, that insurers must possess but may not present in a non-IID manner. Yet, a portion of the considered features is commonly deemed sensitive or exposed during cross-insurer modeling efforts and therefore require special attention. Information linked to historical claims, drivers' record, and socio-demographic data are usually considered sensitive user features, while operational images associated to vehicles are perceived as sensitive on a data partner perspective.

## 6. Evaluation Methodology

The evaluation methodology addresses the establishment of a procedure to assess the federated learning model in a cross-insurer context, specifically the federated framework for cross-insurer prediction defined. Two main aspects will be evaluated: predictive performance and communication efficiency.

To validate the aforementioned risk prediction modeling in a multisite context without data exchange, a federated solution is developed and compared with conventional isolated predictions and a central approach trained on the union of the insurers' databases. The isolated models adopted by each lender with local data are complemented by a central regressor built on an emulated pan-insurer database. The effectiveness of both the central prediction and the isolated models is demonstrated in a base scenario in which risk data are not distributed across locations. The evaluation comprises a synthesis of metrics commonly utilized in federated learning and comparative tests that go beyond adherence to the framework. The predicted value is thus evaluated in terms of accuracy, bias, and variance since neutrality is a key requirement for risk calculation. A fairness index is also employed, since risk is associated with a social cost. The level of privacy leakage, taking the data locality concept into account, and the amount of communication needed to support model training are further considered.

### Equation 3) Federated training loop and FedAvg update number of communication rounds $N$ ,

- client fraction per round  $C \in (0,1)$ ,
- local epochs  $E$ .

Let global model at round  $t$  be  $w^{(t)}$ .

#### Step 1: Server samples clients

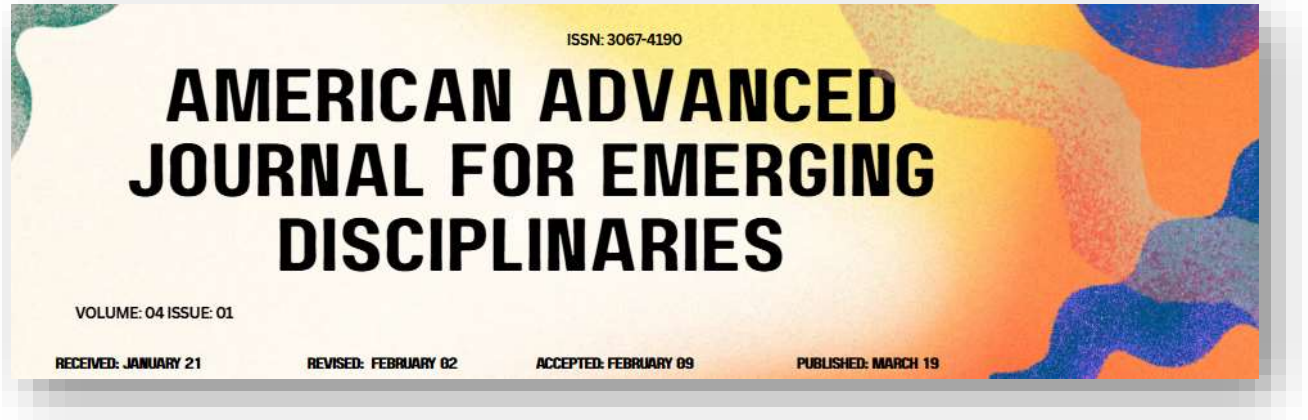
Let  $S_t \subset \{1, \dots, K\}$  be the sampled set, with  $|S_t| \approx CK$ .

#### Step 2: Each sampled client runs local SGD for $E$ epochs

Initialize local model:

$$w_k^{(t,0)} = w^{(t)}$$

For local step  $s = 0, 1, \dots, S - 1$  (where  $S$  depends on epochs  $E$ , batch size  $B$ , and  $n_k$ ):



$$w_k^{(t,s+1)} = w_k^{(t,s)} - \eta g_k^{(t,s)}$$

where a minibatch gradient is

$$g_k^{(t,s)} = \frac{1}{|B_{k,s}|} \sum_{(x,y) \in B_{k,s}} \nabla_w \ell(w_k^{(t,s)}; x, y)$$

After local training, client returns:

$$w_k^{(t+1)} \equiv w_k^{(t,s)}$$

### Step 3: Server aggregates by FedAvg

FedAvg is the weighted average of client models (or equivalently the weighted average of updates):

$$w^{(t+1)} = \sum_{k \in S_t} \frac{n_k}{\sum_{r \in S_t} n_r} w_k^{(t+1)}$$

### 6.1. Dataset Simulation and Benchmarking Scenarios

Both the predictive task and the federated learning (FL) framework supporting it require the cross-insurer scenario to proceed without data sharing. Thus, the potential of FL for collaboration amid privacy restrictions needs to be examined through experimentation that simulates these data silos. An existing real dataset from a single insurer is adopted as a case study, and harmful privacy breaches are prevented via K-Anonymous Framework (K-AF), already benchmarked in the existing literature.

The data of an operatively cross-insurer scenario are simulated to study the effect of data heterogeneity on model performance and explainability. The main insurer data act as the target, while those of other institutions inform the predictive task. A second experimental setting compares a low-data classical approach against a federated learning solution, simulating collaboration without sharing the underlying data. Finally, to represent asymmetries typical of real-world scenarios, the scenarios consider three redistributions of a single institution's data.

### 6.2. Evaluation Metrics

Key factors affecting the efficiency of federated learning in predicting auto insurance risk are:

1. prediction accuracy and reliability;
2. fairness and absence of bias toward certain groups;
3. privacy leakage in the trained global model; and
4. communication cost.

Group fairness constraints should ensure that the global model exhibits similar prediction quality across protected groups, thereby enabling cross-insurer collaboration. Privacy leakage can be assessed using various attack methods from the domain-adaptive federated learning literature. Learning with heterogeneous data improves generalization, but intuitive metrics for these advantages in federated meta-learning remain to be investigated.

## 7. Conclusion

Recent advances in federated learning have enabled computation in a distributed manner while keeping local data at each institution, thereby also alleviating privacy concerns originating from data sharing. An application of federated learning to cross-insurer prediction of car insurance risks from policyholder characteristics has nevertheless been largely uninvestigated. The relevant predictive task, however, is extremely demanding and requires the concentration of data in one central database to be solved with any reasonable accuracy. Thus, a federated learning scheme has been proposed to enable auto risk assessment through collaboration across separate insurers without the need to share raw data.

The proposed framework particularly addresses two issues. First, it defines the adaptive design of the federated learning server for knowledge discovery from distributed, heterogeneous data. Second, it describes the effects of data privacy requirements, Privacy Regulation Compliance by Design, and emerging certification technologies for GDPR regulation in the insurance industry on the federated communication exchange, knowledge aggregation, and cross-insurer collaboration. The employed age- and gender-classified federated knowledge-aggregation scheme adapts

to each individual-local prediction model and preserves customer privacy.

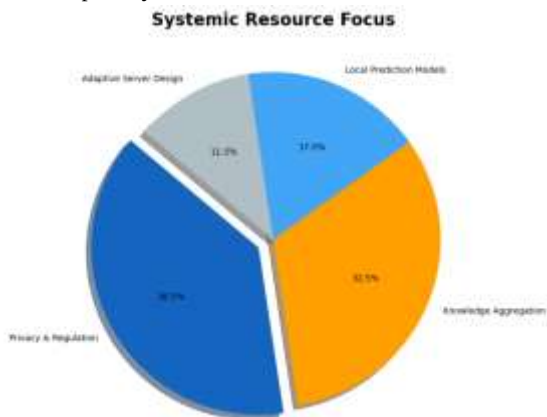


**Fig 4: Systemic Resource Focus**

### 7.1. Final Thoughts and Future Directions

Although trained on non-IID private data, local models achieve promising predictive performance when combined in a federated setting. The absence of data-sharing, including sharing labeled instances, establishes a unique contribution and reduces the risk of privacy breaches. Data controllers wishing to maintain ownership and confidentiality of their data, yet wanting to gain better insights and predictive capability by teaming up with others, can take these aspects into consideration and possibly align legislation with such collaborations. Several research directions remain unexplored: evaluation of different model architectures, alternative predictive tasks, privacy-preserving predictions on federated embedded models, and development of functional federated pipelines from data preprocessing to model predictions. Refining the simulation with real data is another challenge. Efforts could be invested in developing federated pipelines for predictive tasks in other industries (e.g. health or finance), ideally on tasks where institutions would gain advantages from teaming up yet are held back by privacy or regulatory issues.

## 8. References

1. Aledhari, M., Razzak, R., Parizi, R. M., & Saeed, F. (2021). Federated learning: A survey on enabling technologies, protocols, and applications. *IEEE Access*, 9, 140699–140725.
2. Fereidooni, H., Marchal, S., Miettinen, M., Mirhoseini, A., Mollering, H., Nguyen, T. D., Rieger, P., Sadeghi, A. R., Schneider, T., Yalame, H., & Zeitouni, S. (2021). SAFElearn: Secure aggregation for private federated learning. *Proceedings of the 2021 IEEE Security and Privacy Workshops*, 56–62.
3. Yandamuri, U. S., Loganathan, R., Davuluri, P. S. L. N., Rani, P. R. S., Kolla, S. H., & Nagubandi, A. R. (2026). Adaptive Intelligence Networks for Humancentered Enterprise Automation and Governance. In *2026 Third International Conference on Innovations in Cybersecurity and Data Science (ICICDS)* (pp. 678–683). IEEE. 2026 Third International Conference on Innovations in Cybersecurity and Data Science (ICICDS). <https://doi.org/10.1109/icicds70526.2026.11604640>
4. Jeon, B., Ferdous, S. M., Rahman, M. R., & Walid, A. (2021). Privacy-preserving decentralized aggregation for federated learning. *IEEE INFOCOM 2021 Conference on Computer Communications Workshops*, 1–6.
5. Kairouz, P., McMahan, H. B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A. N., Bonawitz, K., Charles, Z., Cormode, G., Cummings, R., D'Oliveira, R. G. L., Eichner, H., El Rouayheb, S., Evans, D., Gardner, M. J., Garrett, Z., Gascón, A., Ghazi, B., Gibbons, P. B., ... Zhao, S. (2021). Advances and open problems in federated learning. *Foundations and Trends in Machine Learning*, 14(1–2), 1–210.
6. Mangalampalli, B. M., Peddi, R. K., Kolla, S. K., Reddy, V. A. R., Mangala, N., & Seenu, A. (2026, June). Explainable Clinical Graph Intelligence Framework for Longitudinal Risk Modeling and Care Pathway Optimization. In *2026 Third International Conference on Innovations in Cybersecurity and Data Science (ICICDS)* (pp. 1–6). IEEE.
7. Zhang, K., Song, X., Zhang, C., & Yu, S. (2022). Challenges and future directions of secure federated

# AMERICAN ADVANCED JOURNAL FOR EMERGING DISCIPLINARIES

VOLUME: 04 ISSUE: 01

RECEIVED: JANUARY 21

REVISED: FEBRUARY 02

ACCEPTED: FEBRUARY 09

PUBLISHED: MARCH 19

- learning: A survey. *Frontiers of Computer Science*, 16, Article 165817.
8. Zheng, Y., Lai, S., Liu, Y., Yuan, X., Yi, X., & Wang, C. (2023). Aggregation service for federated learning: An efficient, secure, and more resilient realization. *IEEE Transactions on Dependable and Secure Computing*, 20(2), 988–1001.
  9. Liu, Z., Lin, H., & Liu, Y. (2023). Long-term privacy-preserving aggregation with user-dynamics for federated learning. *IEEE Transactions on Information Forensics and Security*, 18, 2398–2412.
  10. Banoth, S., Santoshi Kumari, M., Lalitha, P., Deepthi, P., Kumar Peddi, R., & Kavitha, P. (2026). An Efficient Hybrid K-Means and Random Forest-Based Approach for Cloud Malware Detection and Privacy Protection. In *2026 6th International Conference on Intelligent Technologies (CONIT)* (pp. 1–6). IEEE. 2026 6th International Conference on Intelligent Technologies (CONIT).  
<https://doi.org/10.1109/conit69683.2026.11621822>
  11. Eltaras, T. A., Sabry, F., Labda, W., Alzoubi, K., & Malluhi, Q. (2023). Efficient verifiable protocol for privacy-preserving aggregation in federated learning. *IEEE Transactions on Information Forensics and Security*, 18, 2977–2990.
  12. Zhang, Z., Li, J., Yu, S., & Makaya, C. (2023). SAFE Learning: Secure aggregation in federated learning with backdoor detectability. *IEEE Transactions on Information Forensics and Security*, 18, 3289–3304.
  13. Li, K. H., Buarque de Gusmão, P. P., Beutel, D. J., & Lane, N. D. (2022). Secure aggregation for federated learning in Flower. *arXiv preprint arXiv:2205.06117*.
  14. Mahadevan, S. (2026). Governed Serverless Automation Ecosystem for AI-Driven Enterprise Integration and Sustainable Cloud Operations. *International Journal of Computer Information Systems and Industrial Management Applications*, 18(16s), 1561-1570.
  15. Sun, L., Tan, Y., & Wang, J. (2022). A survey on federated learning systems: Vision, applications, and future directions. *IEEE Transactions on Artificial Intelligence*, 3(5), 789–807.
  16. Yang, Q., Liu, Y., Chen, T., & Tong, Y. (2022). Federated machine learning: Concept and applications. *ACM Transactions on Intelligent Systems and Technology*, 13(2), Article 12.
  17. Meng, S., Gao, Y., & Huang, Y. (2022). Actuarial intelligence in auto insurance: Claim frequency modeling with driving behavior features and improved boosted trees. *Insurance: Mathematics and Economics*, 106, 115–127.
  18. Rafi, T. H., Noor, F. A., Hussain, T., & Chae, D.-K. (2023). Fairness and privacy preserving in federated learning: A survey. *Information Fusion*, 102198.
  19. Piyush Kumar Pareek. (2026). Human-Centric Machine Learning Frameworks for Scalable Software Quality Prediction. *Journal of Intelligent Decision Making and Information Science*, 3(5s), 1525–1543.  
<https://doi.org/10.59543/jidmis.v3.1284>
  20. Li, H.-J., Li, J.-Q., & colleagues. (2023). Driving risk prevention in usage-based insurance services based on interpretable machine learning and telematics data. *Decision Support Systems*, 172, 113985.
  21. Strada, S., Costantini, E., Formentin, S., & Savaresi, S. M. (2025). Machine learning based car accident risk prediction for usage-based insurance. *Intelligent Data Analysis*, 29(1), 1–20.
  22. Yao, Y., Sun, Z., & Zhang, Y. (2021). A survey on security and privacy of federated learning. *Future Generation Computer Systems*, 115, 619–640.
  23. Mothukuri, V., Parizi, R. M., Pouriyeh, S., Huang, Y., Dehghantanha, A., & Srivastava, G. (2021). A survey on security and privacy of federated learning. *Future Generation Computer Systems*, 115, 619–640.
  24. Vakkalagadda, T., & Rajamanickam, V. (2026). Agentic AI Architectures for Next-Generation Investment Advisory and Portfolio Intelligence. *International Journal of Computer Information Systems and Industrial Management Applications*, 18(9s), 1206-1219.
  25. Zhang, X., Li, Y., Cui, Z., Cui, Y., & Zhang, J. (2022). A survey on federated learning: Challenges and

# AMERICAN ADVANCED JOURNAL FOR EMERGING DISCIPLINARIES

VOLUME: 04 ISSUE: 01

RECEIVED: JANUARY 21

REVISED: FEBRUARY 02

ACCEPTED: FEBRUARY 09

PUBLISHED: MARCH 19

- applications. *International Journal of Machine Learning and Cybernetics*, 13, 1–18.
26. Saha, S., Hota, A., Chattopadhyay, A. K., Nag, A., & Nandi, S. (2024). A multifaceted survey on privacy preservation of federated learning: Progress, challenges, and opportunities. *Artificial Intelligence Review*, 57, Article 184.
  27. Zhao, J., Bagchi, S., Avestimehr, S., Chan, K., Chaterji, S., Dimitriadis, D., Li, J., Li, N., Nourian, A., & Roth, H. (2025). The federation strikes back: A survey of federated learning privacy attacks, defenses, applications, and policy landscape. *ACM Computing Surveys*, 57(9), Article 230.
  28. Loganathan, R. (2026). SaaS Entitlement Governance: A Reproducible Model for Detecting and Reclaiming Over-Provisioned Access at Enterprise Scale. *Journal of Intelligent Decision Making and Information Science*, 3(7s), 2519-2530.
  29. Li, J., Liu, Z., & Wang, H. (2023). Privacy-preserving federated learning via functional encryption, revisited. *IEEE Transactions on Information Forensics and Security*, 18, 1855–1869.
  30. Xie, H., Wang, Y., Ding, Y., & Yang, C. (2023). Verifiable federated learning with privacy-preserving data aggregation for consumer electronics. *IEEE Transactions on Consumer Electronics*.
  31. Bhavani, B. D., SR, S., Loganathan, R., & Nagaraj, S. (2026, April). Evolutionary Gravitational Neocognitron Neural Network, Snow Leopard Optimization and Deep Graph Reinforcement Learning for Routing Protocol in WSN. In *2026 2nd International Conference on Intelligent Systems and Computational Networks (ICISCN)* (pp. 1-8). IEEE.
  32. Hahn, C., Kim, H., Kim, M., & Hur, J. (2023). VerSA: Verifiable secure aggregation for cross-device federated learning. *IEEE Transactions on Dependable and Secure Computing*, 20(1), 36–52.
  33. Davuluri, P. S. L. (2023). AI-Augmented Sanctions Screening: Enhancing Accuracy and Latency in Real Time Compliance Systems. *AI-Augmented Sanctions Screening: Enhancing Accuracy and Latency in Real Time Compliance Systems* (December 15, 2023).
  34. Jiang, Y., Ma, B., Wang, X., Yu, G., Sun, C., & Liu, R. P. (2023). A secure aggregation for federated learning on long-tailed data. *IEEE Internet of Things Journal*.
  35. Zhao, Y., Chen, J., Wu, D., Teng, J., & Yu, S. (2022). Multi-task network for privacy-preserving federated learning. *IEEE Transactions on Dependable and Secure Computing*.
  36. Ganesh, S. K., Subbareddy, K., Gopi, A., Davuluri, P. N., Mannar, B. R., & Karnawat, A. T. (2026, June). Fraudulent Credit Card Transaction Detection Using a Hybrid Ensemble-Anomaly Detection Framework. In *2026 6th International Conference on Intelligent Technologies (CONIT)* (pp. 1-6). IEEE.
  37. Yang, W., Zhang, Y., Ye, K., Li, L., & Xu, W. (2021). FFD: A federated learning system for fraud detection. *IEEE Transactions on Dependable and Secure Computing*.
  38. Ubale, A. (2025). Privacy-preserving federated learning frameworks for telematics data in auto insurance analytics. *International Journal of Artificial Intelligence, Data Science, and Machine Learning*, 6(4), 41–48.
  39. Piyush Kumar Pareek. (2026). Self-Evolving Analytics Pipelines for Reliable AI-Augmented Software Systems. *Journal of Intelligent Decision Making and Information Science*, 3(5s), 1558–1578. <https://doi.org/10.59543/jidmis.v3.1286>
  40. Davis, J., Smith, J., & Williams, R. (2026). Design of privacy-preserving federated learning models for auto insurance telematics. *ScienceOpen Preprints*.
  41. Anonymous. (2026). Secure aggregation with verifiability and robustness for privacy-preserving federated learning. *Knowledge-Based Systems*.