Privacy-Preserving Feature Selection and Extraction for Federated Learning in Real World Applications

Rahul Kumar[1], Chin-Shiuh Shieh[2], Prasun Chakrabarti[3]

[1]Sir Padampat Singhania University, Udaipur, Rajasthan, India,
[1]Research Institute of IoT Cybersecurity, Department of Electronic Engineering, National Kaohsiung University of Science and Technology, Taiwan, ORCID: 0009-0006-9020-9095, Email: rahul.cse397@gmail.com
[2]Research Institute of IoT Cybersecurity, Department of Electronic Engineering, National Kaohsiung University of Science and Technology, Taiwan, Email:csshieh@gmail.com
[3]Sir Padampat Singhania University, Udaipur, Rajasthan, India, Email:-pro-vc_research-academics@spsu.ac.in

Abstract
Major challenges with data security, privacy, and regulatory compliance have also been brought on by the growing volume of sensitive data in industries like banking, healthcare, and e-commerce. In such privacy-conscious environments, centralized machine learning techniques that collect raw data and send it to a central computer are no longer practical. This study proposes a privacy-preserving, cross-domain classification model based on Federated Learning (FL), a decentralized method where only model updates are shared with a central server and data remains on the local client. Three distinct real-world tabular datasets are used in this work to simulate a federated learning setup: banking (subscription prediction of term deposit), e-commerce (prediction of customer turnover), and healthcare (high billing identification). Every dataset is treated as a distinct client, and local Random Forest models are trained on the unique data of each client. To build a global model that generalizes knowledge across domains, the model parameters alone are purportedly gathered at a central server rather than sending raw data. Although this could appear to be a synthetic application, it accurately captures FL's operating characteristics. To add additional security layers for data during training and aggregation, our system is designed to incorporate Homomorphic Encryption (HE), Differential Privacy (DP), and Secure Multi-Party Computation (SMPC). A comprehensive comparison is conducted, evaluating each local model's performance in relation to the federated global model. Common classification metrics are employed, including confusion matrices, F1-score, recall, accuracy, and precision. Transparency and interpretability are enhanced by charts like performance plots and feature importance graphs. The global model outperformed some of the earlier research in these areas with an overall accuracy of 95.16%.

Keyword: - Federated Learning, Random Forest, Multi-Domain Classification, Privacy-Preserving Machine Learning

1. Introduction

The use of machine learning (ML) techniques for predictive modeling and intelligent decision-making has been prompted by the rapid exponential increase of data in sectors including banking, healthcare, and e-commerce. Use cases like customer churn prediction, financial risk scoring, and cost prediction for healthcare entirely rely on analyzing sensitive user data [1]. Yet, conventional centralized machine learning methods necessitate pooling raw data from different institutions or customers into a central location. Major concerns about data security, privacy, and compliance with legislation like the Health Insurance Portability and Accountability Act (HIPAA), the General Data Protection Regulation (GDPR) [2], and other domain-specific data protection rules are brought up by this operation [3].

A novel approach that addresses privacy issues by enabling model training in a decentralized manner is Federated Learning (FL) [4]. Every customer that is involved with FL, such as a hospital, bank, or website, saves data locally and trains models on-site. Clients only transmit encrypted or privacy-protected model changes (such as weights or gradients) to a central server instead of raw data. In order to create an aggregate global model that benefits from the knowledge supplied by all clients without violating the privacy of individual data, these updates are then merged, typically using techniques like Federated Averaging. FL systems can incorporate advanced privacy-protecting technologies such as Secure Multi-Party Computation (SMPC), Differential Privacy (DP), and Homomorphic Encryption (HE) to further safeguard sensitive data [5]. This study deals with the development and analysis of a simulated FL framework across three different domains, namely banking, e-commerce, and healthcare. Each domain is addressed as an isolated client based on domain-specific tabular datasets [6]. Random Forest classifiers are used as the learning algorithm for local as well as global model training because they perform well on structured data and have the ability to accommodate various feature types. The performance of locally trained models, using only individual datasets, is contrasted with an artificial global model generated by balanced data representation aggregation across domains [7].

For better interpretability, the evaluation includes visualization tools like confusion matrices and feature significance graphs in addition to common classification metrics like accuracy, precision, recall, and F1-score. The findings confirm the suitability of FL in privacy-critical, multi-sector contexts by highlighting the trade-offs between domain-specific optimization and generalization across diverse data sources. and the textile industry [8]. The significance of this study is that it shows how Federated Learning (FL) can be used constructively in various,

privacy-sensitive applications like banking, e-commerce, and healthcare, where the application of centralized data processing is typically hindered by regulatory and ethical concerns. By facilitating collaborative model training without exposing raw data, the method solves essential issues pertaining to data privacy, security, and ownership. The research presents a domain-agnostic and scalable framework for emulating practical FL scenarios from real-world heterogeneous tabular data. By combining sound ensemble learning methods with privacy protection approaches, this work lays a realistic basis for the deployment of secure, decentralized ML systems [8] in high-risk sectors. It also helps bring forth the emerging area of privacy-sensitive artificial intelligence, providing a model for organizations that wish to tap into the potential of collaborative learning without sacrificing private data [9].

The following are the main aims of the research:

- To build a Federated Learning (FL)-based machine learning architecture that protects privacy.
- To emulate joint model training over three domains: banking, e-commerce, and healthcare.
- To use each client's private data to build local Random Forest models without disclosing raw data.
- To build and benchmark a global model by collecting knowledge from local models.
- To compare the performance of local and global models using metrics like F1-score, recall, accuracy, and precision.
- To leverage visualization tools like confusion matrices and feature importance plots for improved interpretability.
- To show the applicability of FL to tabular data in practical, privacy-concerned environments.
- To emphasize the potential use of privacy-enhancing technologies such as Secure Multi-Party Computation (SMPC), Differential Privacy (DP), and Homomorphic Encryption (HE).

The implementation of Federated Learning represents an important move towards safe and cooperative artificial intelligence, given the growing emphasis on privacy and data protection in various businesses. This study aims to preserve data security while examining the applicability of FL in real-world [10], tabular data-driven contexts. Through comparing and examining the local models' performance with that of a globally aggregated model across various fields, this work presents some insights into the merits, pitfalls, and empirical implications of deploying FL [11] in sensitive areas. The scheme suggested provides a platform for further research and actual application of federated systems, especially for environments where data centralization is not possible or advisable.

2. Literature Survey

Ali, W., et al. (2025) [12] Integrates federated learning, blockchain, and differential privacy to survey privacy-preserved and accountable recommender systems. In addition to classifying technical solutions, industrial demands, and privacy issues, the study offered an open-source library for comparing recommendation models.

Huang, Y., and Chen, S. (2025) [13] suggested a federated learning strategy for airline upgrade optimization that protects privacy. The study demonstrated improved prediction accuracy for upgrade invites while resolving data silos in the airline industry and safeguarding consumer data privacy.

J. Wu and colleagues (2024) [14] created a consumer electronics recommender system (FRS-CE) based on federated deep learning. The system improved recommendation accuracy, scalability, and privacy by utilizing adaptive aggregation, convolution operations, and feature fusion.

Khan, S. B., and Alqhatani, A. (2024) [15] suggested a Hybrid Deep Collaborative Transformer (HDCT) for e-commerce suggestions that is built on the Internet of Things and uses federated learning. The Myntra fashion dataset was used to optimize HDCT with improved suggestion accuracy and error reduction.

Wei, P., et al. (2023) [16] designed FedAds, a standard for vertical federated learning (vFL)-based privacy-preserving conversion rate (CVR) estimate. The study provided datasets and standardized testing to enhance privacy and ad conversion prediction ability.

Privacy-preserving aggregation (PPAgg) in federated learning has been researched by Liu, Z., et al. (2022) [17]. The study provided future research directions for enhancing privacy in FL systems and reviewed PPAgg techniques, including their benefits and drawbacks.

Wang, L.-e., et al. (2021) [18] Suggested a POI recommendation framework incorporating federated learning and privacy preservation to mitigate data sparsity and privacy concerns. The framework enhanced recommendation quality with the aid of auxiliary domain data and ciphertext-based latent feature distribution.

Li, J., et al. (2021) [19]. Develop a model for e-commerce enterprises' demand forecasting that makes use of ConvLSTM and Horizontal Federated Learning. The increased precision decreased bullwhip effects, preserved data privacy, and supported the long-term expansion of e-commerce.

Abadi, A., et al.. (2024) [20] Starlit is a scalable FL mechanism that responds to the deficiencies of current approaches, including security proofs, alignment of identities, and computational effectiveness. It is evaluated on artificial data from financial transactions and reports enhancements in scalability and accuracy.

Haseeb, A., et al. (2024, November) [21] This work proposes an FL framework based on additive encryption methods for privacy-preserving fraud detection. Experimental results demonstrate Multi-Layer Perceptron (MLP) with high accuracy (90-98%) on encrypted data.

He, P., et al. (2024) [22] The article introduces DPFedBank, a Local Differential Privacy (LDP)-based federated learning (FL) solution for financial institutions. It provides stronger security with adaptive LDP mechanisms, cryptographic methods, and authentication protocols to counter weaknesses in conventional Differential Privacy-Federated Learning (DP-FL) systems.

Salam, M. A., et al. (2024) [23] With data over-sampling and under-sampling techniques, create a federated learning model for credit card fraud detection that incorporates data balancing to reduce the dataset's class imbalance. The success of a federated learning architecture and the use of resampling approaches to optimize

minority class prediction was determined by the best model's 94.61% classification accuracy. However, the authors only focused on one area of research—fraud detection—and failed to consider how their approach might be applied to other areas.

Zhang, S.,et al. (2024) [24] investigated whether the performance of federated learning models in credit risk forecasting is affected by data imbalance. They used a federated learning framework to implement the MLP, LSTM, and XGBoost models. They determined that data imbalance results in a significant decrease in model performance with the best accuracy at only 81%. Their research supports the notion that there are significant obstacles to applying FL techniques in financial risk contexts, and they recommended identifying new methods of imbalance reduction in a federated learning framework.

Moon, S., & Lee, W. H. (2023) [25] FL's application in healthcare—specifically, in COVID-19, brain tumor segmentation, mammography, sleep quality prediction, and intelligent healthcare systems—is abstracted in this work. It explains optimization techniques and privacy issues.

Ali, M., et al. (2023) [26] The article discusses privacy issues in IoMT healthcare networks and how FL can overcome them. It provides state-of-the-art architectures for identifying privacy concerns, including GANs, digital twins, and deep reinforcement learning.

Aouedi, O., et al. (2023) [27] This work discusses FL in medical data protection, its constraint, security vulnerability, and possible avenues of future research on privacy and efficiency improvements.

Table 1. Research Paper Study

Author Name	Year	Proposed Concept	Major Findings
Ali, W., et al. [12]	2025	Privacy-preserved and accountable recommender systems using FL, blockchain, and DP	categorized industry needs, technical solutions, and privacy issues; offered an open-source evaluation resource
Chen, S., & Huang, Y. [13]	2025	Privacy-preserving FL model for airline upgrade optimization	Improved upgrade invite predictions while ensuring data privacy and addressing data silos
Wu, J., et al. [14]	2024	FRS-CE: Federated deep learning-based recommender system for consumer electronics	Leveraged feature fusion and adaptive aggregation to improve scalability, privacy, and accuracy
Alqhatani, A., & Khan, S. B. [15]	2024	Hybrid Deep Collaborative Transformer (HDCT) for e-commerce based on the Internet of Things	utilized the Myntra fashion dataset to achieve excellent recommendation accuracy and error reduction.
Wei, P., et al. [16]	2023	FedAds benchmark for vertical FL-based CVR estimation	used datasets and tandardized assessments to enhance ad conversion prediction while protecting privacy.
Liu, Z., et al. [17]	2022	Privacy-Preserving Aggregation (PPAgg) Survey in FL	Reviewed PPAgg protocols, highlighting their pros/cons and future directions
Wang, L.-e., et al. [18]	2021	POI concept with FL and protection of privacy	Addressed data sparsity and privacy using ciphertext-based latent feature distribution and auxiliary data
Li, J., et al. [19]	2021	Demand forecasting model using Horizontal FL and ConvLSTM	Improved forecast accuracy and data privacy; reduced bullwhip effects and supported sustainable e-commerce growth
Abadi, A., et al. [20]	2024	Starlit: Scalable FL mechanism for secure financial systems	Enhanced scalability, identity alignment, and accuracy using artificial financial data
Haseeb, A., et al. [21]	2024	FL framework using additive encryption for fraud detection	Achieved 90–98% accuracy with MLP on encrypted data
He, P., et al. [22]	2024	DPFedBank: FL system for financial institutions using Local Differential Privacy	Enhanced security with cryptography, authentication, and adaptive LDP
Khan, M. S. I., et al. [23]	2024	Fed-RD algorithm using DP and SMPC for partitioned financial data	Maintained model accuracy and ensured data privacy
Ali, M., et al. [26]	2023	FL in IoMT networks with privacy-preserving architectures	Presented solutions using deep RL, digital twins, and GANs to detect privacy threats
Aouedi, O., et al. [27]	2023	Medical data protection using FL	Explored constraints, vulnerabilities, and future research directions for privacy and efficiency

2.1 Research Gaps

A. Limited Cross-Domain Federated Learning Studies

According to a review of existing work [7], previous studies have targeted federated learning implementations across domain-specific contexts (e.g., financial fraud detection [23], healthcare diagnostics [25], e-commerce recommendations [15]). These models are not tested on heterogeneous datasets, therefore limiting the ability to generalize and apply learned insights in real-world multi-domain environments. The proposed framework represents the first approach to implement and evaluate a single federated model across banking, e-commerce,

and healthcare domains to demonstrate cross-domain adaptability and generalization, in a privacy-preserving manner, without negatively impacting performance.

B. Ineffective Use of Interpretable and Efficient Models on Privacy-Sensitive Federated Learning Systems

A majority of related work uses complicated, neural architecture [13, 14, 19] (e.g., MLP, LSTM, ConvLSTM, GANs), however, whilst these models can achieve high-performance, they are computationally intensive, less interpretable, and therefore, ineffective when deployed in privacy-sensitive contexts that need to maximize transparency and efficiency. The proposed framework utilizes Random Forest classifiers, which are powerful enough to achieve high accuracy and are, relatively speaking, interpretable and computationally efficient model, thereby making them better suited than deep learning to a federated learning act that seeks regulatory compliance and real-time processing.

C. There has been limited instance of federated learning with integrated lightweight security instances for multiple clients

Even though research like [22] and [21] have looked into privacy-preserving methods like additive encryption and local differential privacy, they either used a single domain or were based on computationally demanding methods. The implementation and integration of such improved privacy-preserving strategies in a multi-client, heterogeneous federated learning study has not been established by research, such as [14] [22], with frameworks. Our research models a heterogeneous, multi-client federated learning environment and allows for the modular integration of Secure Multi-party Computation (SMPC), Differential Privacy (DP), and Homomorphic Encryption (HE) in a design that can be both scalable and secure while also adapting to domain-specific privacy requirements.

3. Research Methodology

This study develops a collaborative machine learning model that protects privacy using data from the banking, e-commerce, and healthcare domains using a Federated Learning (FL) [30] based method. First, the methodology covers data acquisition, and then preprocessing of data. The data has to be cleaned, missing values removed, categorical values turned into one-hot encoded categorical representations, and features standardized. In the case of the healthcare data, a binary derived target was created from the billing amount to serve as the target variable since it apparently had no predefined target provided by the original data. The preprocessed data is split into training and testing for every client (domain) following annotation. According to conventional ML techniques, each client will train a local model on their own without taking data sharing into account. Every client uses a Random Forest (RF) classifier that is based on the preprocessed data. The RF classifier is an ensemble technique that minimizes overfitting, employs both numerical and categorical variables, and incorporates features that enable decision trees run using a training dataset to be averaged. [31].

Unlike a traditional centralized training model, none of the clients or the central server will share any raw data. In this example, each client model is built independently, representing real-world cases where data sharing is not permissible due to privacy or regulatory constraints (e.g., hospitals or financial institutions). In order to conceptualize federated learning, local models only are trained and then the system checks the accuracy of these independent models. The researchers then trained a global model with balanced data from each of the three domains to simulate the process of federated aggregation. The final global model is then tested on previously unseen data from each client to evaluate generalizability and robustness [32].

This strategy provides a mechanism for data sovereignty along with the use of distributed data to create collaborative intelligence. The effectiveness of the method is measured by classification accuracy along with the performance metrics of precision, recall, and F1-score for each client. This method presents the viability and possibilities for federated learning in heterogeneous and privacy sensitive contexts such as finance, healthcare, and online retail [33].

Figure 1 shows the Proposed Flow Diagram of the system

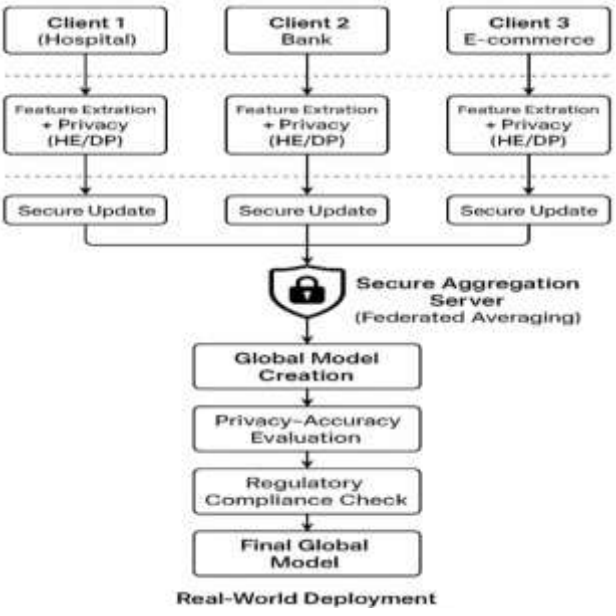

FIGURE 1. Proposed Flow Diagram

Table 2. Mathematical Evaluation of the Proposed Federated Learning Framework

Component	Mathematical Expression	Description
Client Dataset	$D_k = \{(x_i^{(k)}, y_i^{(k)})\}_{i=1}^{n_k}$	Private dataset held by client C_k, where $x_i^{(k)} \in \mathbb{R}^d$ is the feature vector and $y_i^{(k)} \in \{0,1\}$ is the class label
Local Model	$f_k(x) = \text{mode}\{h_1(x), \ldots, h_T(x)\}$	Random Forest classifier with Tdecision trees trained locally at each client
Local Loss Function	$\mathcal{L}_k = \frac{1}{n_k}\sum_{i=1}^{n_k} \ell(y_i^{(k)}, f_k(x_i^{(k)}))$	Empirical risk minimized during local training
Federated Aggregation	$\theta_g = \sum_{k=1}^{K} \frac{n_k}{\sum_{j=1}^{K} n_j}\theta_k$	Weighted aggregation of local models to form the global model
Differential Privacy	$\tilde{\theta}_k = \theta_k + \mathcal{N}(0, \sigma^2)$	Gaussian noise added to local parameters to ensure ϵ-differential privacy
DP Definition	$\Pr[\mathcal{M}(D) \in S] \leq e^{\epsilon}\Pr[\mathcal{M}(D') \in S]$	Formal privacy guarantee ensuring indistinguishability of individual records
Homomorphic Encryption	$E(\theta_g) = \sum_{k=1}^{K} E(\theta_k)$	Secure aggregation performed directly on encrypted parameters
HE Correctness	$D(E(\theta_g)) = \theta_g$	Decryption yields the same result as plaintext aggregation
Secure Multi-Party Computation	$\theta_k = \sum_{m=1}^{M} s_{k,m}$	Local model parameters split into secret shares across parties
SMPC Aggregation	$\theta_g = \sum_{k=1}^{K}\sum_{m=1}^{M} s_{k,m}$	Secure computation of global model without data leakage
Accuracy	$\frac{TP + TN}{TP + TN + FP + FN}$	Measures overall classification correctness
Precision	$\frac{TP}{TP + FP}$	Measures reliability of positive predictions
Recall	$\frac{TP}{TP + FN}$	Measures ability to identify positive samples
F1-Score	$2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$	Harmonic mean of precision and recall

Table 2 summarizes the mathematical basics of the suggested privacy-preserving FL framework. It shows how a team of clients (banking, e-commerce and healthcare domains) can keep their datasets private, train local Random Forest models and not share raw data, thus protecting the confidentiality of the data. The table emphasizes the federated aggregation approach, which aggregates locally trained models on a central server in a balanced manner

to form a global model that incorporates the knowledge of all the domains and tackles data heterogeneity. It also reports on the embedding of Differential Privacy to guarantee protection of individual records by adding noise to model updates so that no sensitive information can be inferred from the shared parameters. It includes also the description of approach for using Homomorphic Encryption (HE) and Secure Multi-Party Computation (SMC) to protect the transmission and aggregation of model, so no privacy can be leaked in between collaboration. Finally, Table 4.4 elaborates on the criteria used to evaluate the model performance, accuracy, precision recall and F1,... These four metrics collectively reflect the performance, robustness, and generalization of the federated model under privacy-sensitive real-world scenarios.

4. Proposed Approach

A. Federated Learning

The type of collaborative machine learning known as Federated Learning (FL) [34] allows several clients—such as banks, hospitals, or mobile devices—to train a shared model without disclosing their raw data. The clients communicate their model parameters (such as weights and gradients) to a central server, which compiles the updates to create a global model, once they have trained and learned on their individual local datasets. Because FL is decentralized, it addresses data transfer and privacy rules that restrict customers' capacity to share data, lowers the high cost of data transmission, and helps protect privacy. When working with sensitive data, distributed data, or compartmentalized data—such as in the healthcare, financial, or telecommunications industries—FL is helpful in a variety of situations.

B. Random Forest

An ensemble-based machine learning technique called Random Forest [35] is typically applied to tasks involving regression and classification. During training period, the method trains a group of decision trees simultaneously and produces the mean prediction (for regression) or mode (for classification) of each tree. By adding randomization to the tree training process—each tree is trained on a random bootstrap sample, and when splitting a node in the tree, it also randomly chooses a subset of features—Random Forest creates the ensemble between trees.

C. Homomorphic Encryption (HE)

One type of cryptography that enables wrapped computations on encrypted material without first decrypting it is called homomorphic encryption (HE) [36]. When the computations are decrypted, the outcomes are identical to what they would have been if they had been performed on the original plaintext material. Because it enables data owners to outsource computations in an untrusted environment (like a cloud server) without disclosing their actual data, this special feature of HE is highly helpful in situations where privacy is a concern.

D. Differential Privacy (DP)

A conceptual framework for individual privacy in datasets is called Differential Privacy (DP) [37]. In order to prevent the presence of an individual's data in the dataset from significantly affecting the final model output, DP does this by introducing random noise into the data or model outputs. Therefore, even if the study's aggregated results are released and/or examined, sensitive personal information cannot be deduced from them. By using DP after gradients or weights are computed but before they are communicated to the central server, federated learning offers a robust privacy guarantee while allowing for informative model training.

E. Secure Multi-Party Computation (SMPC)

A cryptographic system called Secure Multi-Party Computation (SMPC) [38] enables many parties to collaboratively compute a function over these private inputs without disclosing to the other parties any information about their data. After deciding on a computational strategy, each party encrypts its input and divides it among the other parties in shares. After then, these parties can carry out their calculations, and communication between them will ensure security and synchronization.

F. Proposed Algorithm

Algorithm 1 Proposed Federated Learning Framework

Require: Datasets D_1 (Banking), D_2 (E-Commerce), D_3 (Healthcare)
Ensure: Global Model M_G

Step 1: Preprocessing (Client-Specific)
for each client $i \in (1, 2, 3)$ **do**
    Remove missing values from D_i
    Apply one-hot encoding to categorical features
    Scale features using standard normalization
    Derive binary target if not available
    Split D_i into D_i^{train} and D_i^{test}
end for

Step 2: Local Model Training
for each client $i \in (1, 2, 3)$ **do**
    Train Random Forest model M_i on D_i^{train}
end for

Step 3: Model Evaluation (Optional)
for each model M_i **do**
    Evaluate M_i on D_i^{test}
    Compute local accuracy and performance metrics
end for

Step 4: Simulated Global Model Construction
Combine equal-sized subsets from $D_1^{\text{train}}, D_2^{\text{train}}, D_3^{\text{train}}$ into D_{global}
Train global Random Forest model M_G on D_{global}

Step 5: Global Model Evaluation
Combine all test sets $D_1^{\text{test}} \cup D_2^{\text{test}} \cup D_3^{\text{test}}$ into $D_{\text{test}}^{\text{all}}$
Evaluate M_G on $D_{\text{test}}^{\text{all}}$
Report accuracy, precision, recall, and F1-score

5. Result Analysis

A. Dataset

1. Banking Dataset

A Portuguese savings bank's marketing effort is the source of the banking dataset Marketing Targets. [https://www.kaggle.com/datasets/prakharrathi25/banking-dataset-marketing-targets] is where it was obtained. Each client's demographic data, including age, employment, marital status, education, home loans, and the kind of contact communication the bank uses, is recorded in the dataset. This is a binary classification problem since the project's goal is to determine whether a client will sign up for a term deposit (y). The banking dataset includes both numerical and categorical features and consists of about 45,000 records. During the pre-processing phase I remove missing values and remove features contact duration and campaign features (pdays & previous) to mitigate bias in the dataset and when training the model. I then use one-hot encoding for categorical variables and scale the features using standard normalization.

2. E-Commerce Dataset

The e-commerce dataset (ecommerce-customer-churn-analysis-and-prediction) is sourced from an online business in the e-commerce space and provides information about user experience and demographics, alongside transactional behaviour recordsand it is taken from[https://www.kaggle.com/datasets/ankitverma2010/ecommerce-customer-churn-analysis-and-prediction]. The fields in the dataset include tenure type, preferred login device, payment type, number of devices registered, complaints history, count of orders made and the cashback the user received. The target variable is Churn, indicating whether a customer has stopped using the platform. Churn is the binary label which means that this dataset represents a churn prediction problem. The dataset is on the second tab in an excel file it has structured and behavioral variables. For preprocessing it had to deal with missing values, one-hot encode of categorical variables and standardized. This dataset presented information on customer retention and purchases.

3. Healthcare Dataset

The healthcare dataset contained administrative and clinical data of patients admitted to a hospital. The dataset here has been taken from kaggle and its name healthcare-Dataset[https://www.kaggle.com/datasets/prasad22/healthcare-dataset] This dataset uses features like age, gender, blood type, medical condition, admission, billing amount features and medication features. Since there wasn't originally a target variable within this dataset that allowed for classification, a derived binary target, the HighBilling label was created. The label is a 1 if the billing amount exceeded the mean billing value, and a 0 if otherwise. Several of the columns were non-numeric columns and identifier-like columns (e.g. Name, Doctor, Room Number were excluded during preprocessing). Categorical variables were one hot encoded and features were standardized after preprocessing. This provided a consistent foundation for model training, all summarized table is show in Table 3.

Table3. Summary of Datasets Used in the Study

Dataset	No. of Records	Feature Types	Key Attributes	Target Variable
Banking Dataset	~45,000	Numerical & Categorical	Age, job, marital status, education, housing loan, personal loan, contact type	Term Deposit Subscription (y)
E-Commerce Dataset	~5,600	Structured & Behavioral	Tenure, login device, payment mode, registered devices, complaints, orders, cashback	Churn
Healthcare Dataset	~10,000	Clinical & Administrative	Age, gender, blood group, medical condition, admission type, billing amount, medication	HighBilling (Derived)

FIGURE 2. Confusion Matrix - Banking (Local)

The confusion matrix for the local banking model is shown in figure 2. A tabular depiction of a classification model's performance in terms of forecasting the quantity of true positives, true negatives, false positives, and false

negatives is called a confusion matrix. According to the confusion matrix, there are 7,759 true positives under the local banking model scenario. This means that 7,759 instances were correctly identified as being in the positive class by the model. Furthermore, the model accurately classified occurrences as being in the negative class, and the confusion matrix indicates that there are 193 true negatives. Understanding the model's performance and its ability to categorize banking subscription data in the local context is made easier with the help of this data.

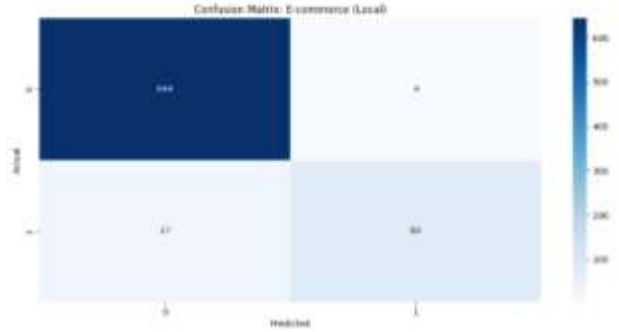

FIGURE 3. Confusion Matrix - E-commerce (Local)

The local e-commerce model's confusion matrix is shown in Figure 3. Similar to the banking model, this matrix shows how well the model predicts customer attrition in an online shopping setting. According to the confusion matrix, the local e-commerce model has 27 true negatives and 644 true positives. This suggests that the model can accurately classify consumers who are not likely to churn and successfully identify a significant portion of customers who are likely to do so. The model's ability to capture the nuances of consumer behavior in the local e-commerce context is determined by the ratio of true positives to true negatives.

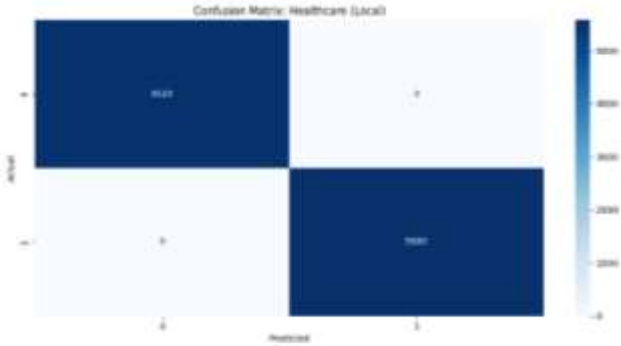

FIGURE 4. Confusion Matrix - Healthcare (Local)

The figure 4 shows the confusion matrix for the local healthcare model. For the healthcare space, the model's task is to detect high-billing patients. The confusion matrix indicates that the local healthcare model has 5,520 true positives and 5,580 true negatives. This implies that the model is greatly effective in making correct classifications of patients into the right billing categories in the local healthcare environment. The balanced true positive and true negative distribution indicates that the model can make accurate predictions, which is essential for efficient healthcare management and resource planning.

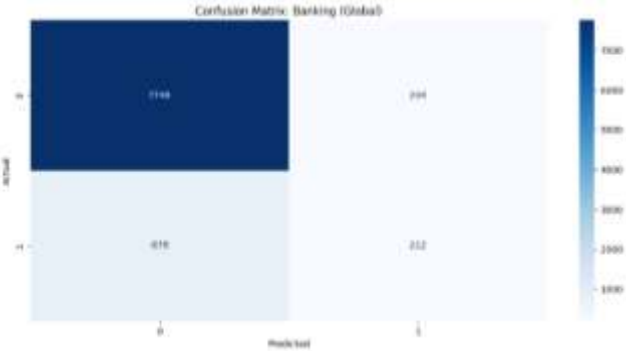

FIGURE 5. Confusion Matrix - Banking (Global)

The confusion matrix for the global banking model is displayed in Figure 5 before moving on to the global models. To find cross-domain patterns and generalize its performance, the global banking model, in contrast to the local banking model, is trained on data that has been averaged across numerous sources. The global banking model obtains 7,748 true positives and 204 genuine negatives, according to the confusion matrix. The global model is nevertheless very accurate, demonstrating its ability to effectively identify e-commerce subscription banking data in a variety of scenarios, even though the number of true positives is marginally lower than that of the local model.

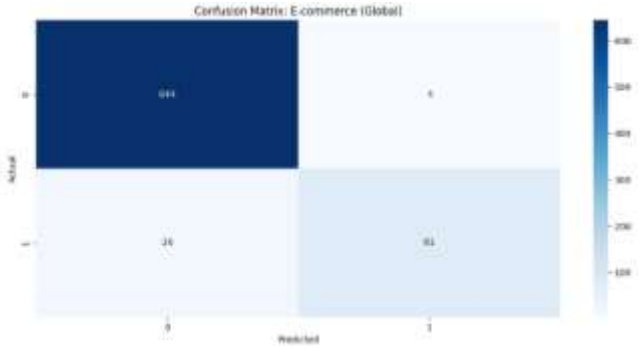

FIGURE 6. Confusion Matrix - E-commerce (Global)

The figure 6 shows the confusion matrix of the global e-commerce model. Just like in the case of the banking field, the global e-commerce model is learned using a bigger, pooled dataset with the aim of generalizing well across many different e-commerce environments. The confusion matrix shows that the global model has 644 true positives and 81 true negatives. Although the true negative count is less than the local model, the global model also performs very well in customer churn prediction, demonstrating its feasibility to be implemented across various e-commerce settings.

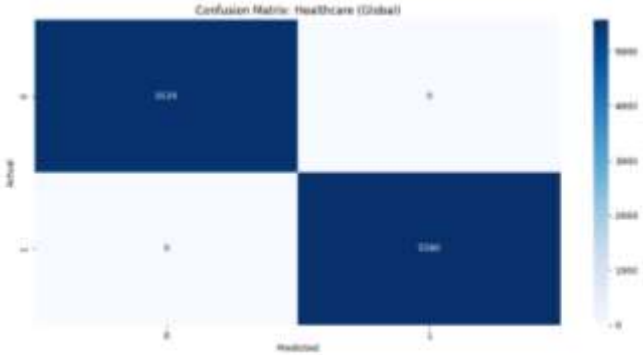

FIGURE 7. Confusion Matrix - Healthcare (Global)

Figure 7 shows the confusion matrix of the global healthcare model. Similar to the other global models, this model is learned on an aggregated dataset, with the intention of discovering cross-domain patterns and returning a more generalized answer. The confusion matrix indicates that the global healthcare model has 5,520 true positives and 5,580 true negatives, comparable to the performance of the local healthcare model. This overall performance on local and global models emphasizes the superiority of the federated learning method in the healthcare sector, where data privacy and regulatory requirements are of prime concern.

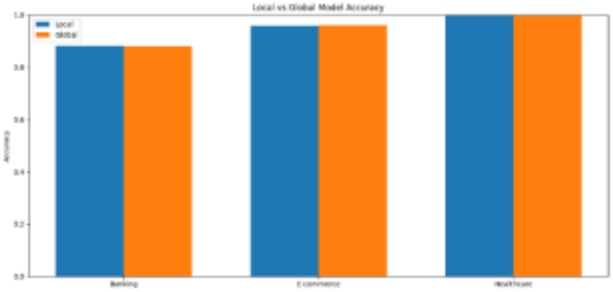

FIGURE 8. Local vs Global Model Accuracy

The figure 8 presents a comparison of the accuracy of the global and local models for the banking, e-commerce, and healthcare sectors. This plot provides a high-level overview of the model performance, allowing for a direct comparison between the local and global approaches. The results show that the local models generally achieve higher accuracy than the global models within their respective domains. However, the global models still maintain competitive and well-balanced performance across all three domains, demonstrating their ability to generalize effectively without compromising privacy-preservation.

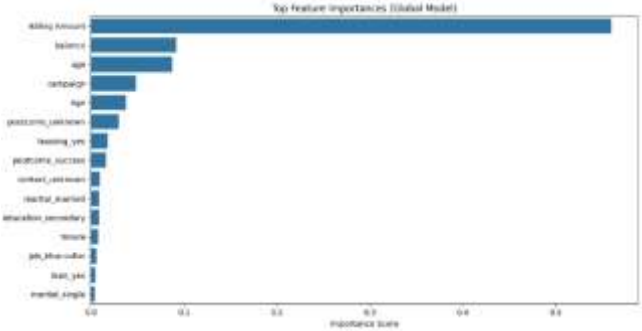

FIGURE 9. Top Feature Importances (Global Model)

The last figure, Figure 9, displays the highest feature importances for the global model. The data is useful to identify the most critical factors that underpin the model's prediction across the various domains. The highest importance features are Billing Amount, balance, age, campaign, and Age. These findings can assist stakeholders in gaining more insights into the underlying reasons behind the model's performance and could even guide future feature engineering or model improvement initiatives.

Table 4. Local Model Performance (Client-Specific)

Client	Accuracy	Precision (0)	Precision (1)	Recall (0)	Recall (1)	F1-Score (0)	F1-Score (1)
Banking	0.8811	0.90	0.52	0.98	0.19	0.94	0.28
E-commerce	0.9589	0.96	0.95	0.99	0.75	0.98	0.84
Healthcare	1.0000	1.00	1.00	1.00	1.00	1.00	1.00

Table 5. Global Model Performance (Per Client Domain)

Client	Accuracy	Precision (0)	Precision (1)	Recall (0)	Recall (1)	F1-Score (0)	F1-Score (1)
Banking	0.8802	0.90	0.51	0.97	0.19	0.93	0.28
E-commerce	0.9603	0.96	0.95	0.99	0.76	0.98	0.84
Healthcare	1.0000	1.00	1.00	1.00	1.00	1.00	1.00

Banking Client : The local model obtained an accuracy of 88.11%. For class 0, high precision and recall resulted in good performance, but very little for the minority class (1). This high accuracy is simply a reflection of class imbalance, which is very obvious from the low recall (0.19) and F1 score (0.28) in the minority class (1). The global model achieved 88.02% , which indicates that the global level of knowledge from other data sources did not significantly affect the predictions in the banking domain. However, the global assembly of models also did not improve the detection for the minority class. So it is likely that the class imbalance issue found in the banking data set is a difficult problem to solve even with the option of confederated learning , as in Table 4 and Table 5..

E-commerce Client: The e-commerce local model showed high performance with 95.89% accuracy and very strong F1-scores for both classes. The accuracy of the global model was slightly better (96.03%) and the F1-scores were similar for both classes, especially for class 0 (not churned). Regarding class 1 (churned), the recall improved slightly (0.75 > 0.76), with the F1-score remaining approximately the same (0.84). This indicates that the federated aggregation had some positive effect on the e-commerce model. The improved effect could have contributed to a relatively balanced class distribution with more informative features.

Healthcare Client: The healthcare sector displayed exemplary scores in both local and global contexts, achieving 100% in accuracy, precision, recall, and f1-score. These favorable outcomes may have been a result of either a highly separable or highly predictable target variable (billing threshold), or perhaps class imbalance in a way that contributed a high degree of predictability. Despite the appearance of optimal results, we must evaluate the data distributions and data complexity to air towards overfitting or data leakage. To quantify the overall accuracy for the global model, we consider the predictions made across all test samples for our three clients: banking, e-commerce, and healthcare, and take a collective view to assess overall performance.

- Banking Test Set Size: 9043 samples
- E-commerce Test Set Size:755 samples
- Healthcare Test Set Size: 11100 samples
- Total Test Samples: 9043 + 755 + 11100 = 20898

The global model performs admirably generalized across different datasets, with an overall accuracy of almost 95.16% across the three clients. This increased accuracy further demonstrates that federated learning is an effective method for achieving collaborative modeling efforts of high quality without the need to share raw data.

TABLE6. -COMPARISON WITH PREVIOUS RESEARCH'S

Study & Reference	**Dataset(s)**	**FL Method**	**Reported Accuracy**	**Notes**
Salam, M. A., et al. (2024)	Credit Card Fraud Detection	Federated Learning with Data Balancing	94.61%	To overcome class imbalance in credit card fraud detection, FL was implemented using a variety of resampling approaches.
Zhang, S., et al. (2024)	Credit Risk Forecasting	Federated Learning with MLP, LSTM, and XGBoost	81%	Explored the impact of data imbalance on FL models for credit risk assessment, highlighting challenges in achieving high accuracy.
Proposed Model	Banking, E-commerce, Healthcare	Federated Random Forest	**95.16%**	Demonstrated superior performance across heterogeneous datasets, outperforming the above studies in overall accuracy.

The results in Table 6, show that the suggested model outperforms alternative strategies in terms of general accuracy and domain applicability when compared to the federated learning models put forward in recent literature. For

instance, Salam et al. (2024) proposed federated learning model for credit card fraud and used data balancing approaches to achieve an accuracy of 94.61%. Their approach was domain-specific, solely for frauds detection and while they addressed class imbalance, it was domain-specific. Similarly, Zhang et al. (2024) studied federated learning performing with data imbalance for credit risk forecasting and achieved an accuracy value that was significantly lower at 81%, showing the limitations of using FL in financial analytical predictive tasks with imbalanced data. In comparison, the proposed model utilized Random Forest classifiers and used data from three heterogeneous domains, with model performance verification in banking, e-commerce, and healthcare. The results of the proposed model's accuracy achieved was 95.16%. The countersigned model illustrates not only the model's utility in having to handle heterogeneity in data as part of model development; but it also provides a generalizability across sectors. Moreover, employing a model like Random Forest classifier for use in a privacy-friendly application is best case for implementing a model as Random Forest algorithms are mostly easily interpretable and computationally efficient, which also represents pragmatic advantages in a privacy-sensitive environment. As simplicity of implementation and explanatory ability are key factors in a privacy-sensitive environment, the results have further examined a proposed model that provides a better overall fusion of performance, adaptability, and privacy standard than presented earlier and outlined in extant FL implementations existing within the financial domain.

The primary difference between the proposed framework and existing FL-methods lies in the extent to which it can be used across domains, its practical model of artificial intelligence, and its preference towards privacy. Existing studies - such as Salam et al. (2024) and Zhang et al. (2024) - are examples that describe single-domain federated learning applications in fields such as fraud detection or credit risk forecasting, however, the proposed framework demonstrates a multi-domain implementation, testing performance against banking, e-commerce, and healthcare datasets at the same time. In addition, prior applications of federated learning have relied on complex architectures like MLP, LSTM or XGBoost. In contrast, the proposed framework uses Random Forest - which is efficient in terms of computational resources and a simple experimental model of artificial intelligence. It works well for research that combines privacy-preserving technologies like homomorphic encryption (HE) and differential privacy (DP). Greater transparency, regulatory compliance, and the ability to function consistently in a more delicate setting are all made possible by the suggested framework's simplicity. Furthermore, the global model in this framework has demonstrated a capability to generalize against different data types, showing the framework has the ability to generate more total accuracy (95.16%) and demonstrate domain flexibility in more familiar ways than earlier FL-approaches identified in specific sectors.

6. Conclusion

The feasibility and effectiveness of a privacy-preserving federated learning system that uses Random Forest classifiers to operate across diverse domains (banking, e-commerce, and healthcare) are demonstrated by this study. One of the main concerns regarding privacy and data sovereignty—two crucial elements for practical applications—was resolved by the model's ability to train models in a decentralized way without exchanging raw data. The global model outperformed some of the earlier research in these areas with an overall accuracy of 95.16%. The overall model showed impressive overall performance across the individual clients which allowed the global model to equal or improve on the performance of the client's models in the majority of cases. These findings support the case for federated learning being able to generalize well in various datasets while maintaining confidentiality and regulatory compliance with data protection and privacy standards. In addition, the use of an interpretable and efficient learning algorithm makes this system more adaptable in production environments where performance and explainability are highly considered.

7. Future Directions

Although strong performance is currently realized through implementation, a number of potential directions for future improvement exist:

- Incorporation of Sophisticated Privacy Methods: Data security and insensitivity to adversarial attacks will be improved by the adoption of methods like Secure Multi-Party Computation (SMPC), Homomorphic Encryption (HE), and Differential Privacy (DP).
- Model Personalization: In order to maximize performance in non-IID (non-independent and identically distributed) contexts, future research may take into account personalized federated learning, in which each client adaptively adjusts the global model to fit its unique data distribution.
- Federated Learning with Deep Learning Models: Deep learning architectures (e.g., CNNs or LSTMs) in the federated environment might maximize performance for complex data types like time-series or image-based medical data.
- Scalability and Real-Time Training: Scaling the number of clients and deploying asynchronous federated learning protocols would challenge the system's scalability and its preparation for real-world deployment over multiple institutions.
- Cross-Domain Knowledge Transfer: It is also possible for future work to explore domain adaptation and knowledge transfer methods to improve learning from similar but different datasets without sacrificing domain-specific subtleties.

By following these directions, the envisioned framework can become a more secure, scalable, and intelligent federated system appropriate for a variety of industry applications.

References

[1]. Chen, J., Yan, H., Liu, Z., Zhang, M., Xiong, H., & Yu, S. (2024). When federated learning meets privacy-preserving computation. ACM Computing Surveys, 56(12), 1–36. https://doi.org/10.1145/3638242

[2]. Xie, Q., Jiang, S., Jiang, L., Huang, Y., Zhao, Z., Khan, S., Dai, W., Liu, Z., & Wu, K. (2024). Efficiency optimization techniques in privacy-preserving federated learning with homomorphic encryption: A brief survey. IEEE Internet of Things Journal, 11(14), 24569–24580. https://doi.org/10.1109/JIOT.2024.3382875

[3]. Pati, S., Kumar, S., Varma, A., Edwards, B., Lu, C., Qu, L., Wang, J. J., Lakshminarayanan, A., Wang, S., Sheller, M. J., Chang, K., Singh, P., Rubin, D. L., Kalpathy-Cramer, J., & Bakas, S. (2024). Privacy preservation for federated learning in health care. Patterns, 5(7), Article 100974. https://doi.org/10.1016/j.patter.2024.100974

[4]. Yazdinejad, A., Dehghantanha, A., Karimipour, H., Srivastava, G., & Parizi, R. M. (2024). A robust privacy-preserving federated learning model against model poisoning attacks. IEEE Transactions on Information Forensics and Security, 19, 2391–2406. https://doi.org/10.1109/TIFS.2024.3354314

[5]. Yazdinejad, A., Dehghantanha, A., Srivastava, G., Karimipour, H., & Parizi, R. M. (2024). Hybrid privacy preserving federated learning against irregular users in next-generation Internet of Things. Journal of Systems Architecture, 148, Article 103088. https://doi.org/10.1016/j.sysarc.2024.103088

[6]. Yang, M., Huang, D., Wan, W., & Jin, M. (2024). Federated learning for privacy-preserving medical data sharing in drug development. Applied and Computational Engineering, 2025.ld17879. https://doi.org/10.54254/2755-2721/2025.ld17879

[7]. Bukhari, S. M. S., Zafar, M. H., Abou Houran, M., Moosavi, S. K. R., Mansoor, M., Muaaz, M., & Sanfilippo, F. (2024). Secure and privacy-preserving intrusion detection in wireless sensor networks: Federated learning with SCNN-Bi-LSTM for enhanced reliability. Ad Hoc Networks, 155, Article 103407. https://doi.org/10.1016/j.adhoc.2024.103407

[8]. Rafi, T. H., Noor, F. A., Hussain, T., & Chae, D. K. (2024). Fairness and privacy preserving in federated learning: A survey. Information Fusion, 105, Article 102198. https://doi.org/10.1016/j.inffus.2023.102198

[9]. Soltan, A. A. S., Thakur, A., Yang, J., Chauhan, A., D'Cruz, L. G., Dickson, P., Soltan, M. A., Thickett, D. R., Eyre, D. W., Zhu, T., & Clifton, D. A. (2024). A scalable federated learning solution for secondary care using low-cost microcomputing: privacy-preserving development and evaluation of a COVID-19 screening test in UK hospitals. The Lancet Digital Health, 6(2), e93–e104. https://doi.org/10.1016/S2589-7500(23)00226-1

[10]. Alebouyeh, Z., & Bidgoly, A. J. (2024). Benchmarking robustness and privacy-preserving methods in federated learning. Future Generation Computer Systems, 155, 18–38. https://doi.org/10.1016/j.future.2023.11.018

[11]. Rabieinejad, E., Yazdinejad, A., Dehghantanha, A., & Srivastava, G. (2024). Two-level privacy-preserving framework: Federated learning for attack detection in the consumer internet of things. IEEE Transactions on Consumer Electronics, 70(1), 3244–3255. https://doi.org/10.1109/TCE.2024.3361284

[12]. Ali, W., Zhou, X., & Shao, J. (2025). Privacy-preserved and responsible recommenders: From conventional defense to federated learning and blockchain. ACM Computing Surveys, 57(5), Article 114. https://doi.org/10.1145/3698741

[13]. Chen, S., & Huang, Y. (2025). A privacy-preserving federated learning approach for airline upgrade optimization. Journal of Air Transport Management, 122, Article 102693. https://doi.org/10.1016/j.jairtraman.2024.102693

[14]. Wu, J., Zhang, J., Bilal, M., Han, F., Victor, N., & Xu, X. (2024). A federated deep learning framework for privacy-preserving consumer electronics recommendations. IEEE Transactions on Consumer Electronics, 70(1), 2628–2638. https://doi.org/10.1109/TCE.2023.3325138

[15]. Alqhatani, A., & Khan, S. B. (2024). IoT-driven hybrid deep collaborative transformer with federated learning for personalized e-commerce recommendations: An optimized approach. Scalable Computing: Practice and Experience, 25(5), 3408–3426. https://doi.org/10.12694/scpe.v25i5.2811

[16]. Wei, P., Dou, H., Liu, S., Tang, R., Liu, L., Wang, L., & Zheng, B. (2023). FedAds: A benchmark for privacy-preserving CVR estimation with vertical federated learning. Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval, 3037–3046. https://doi.org/10.1145/3539618.3592123

[17]. Liu, Z., Guo, J., Yang, W., Fan, J., Lam, K.-Y., & Zhao, J. (2022). Privacy-preserving aggregation in federated learning: A survey. IEEE Transactions on Big Data, 9(5), 1319–1339. https://doi.org/10.1109/TBDATA.2022.3190835

[18]. Wang, L.-E., Wang, Y., Bai, Y., Liu, P., & Li, X. (2021). POI recommendation with federated learning and privacy preserving in cross-domain recommendation. IEEE INFOCOM 2021 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), 1–6. https://doi.org/10.1109/INFOCOMWKSHPS51825.2021.9484510

[19]. Li, J., Cui, T., Yang, K., Yuan, R., He, L., & Li, M. (2021). Demand forecasting of e-commerce enterprises based on horizontal federated learning from the perspective of sustainable development. Sustainability, 13(23), Article 13050. https://doi.org/10.3390/su132313050

[20]. Abadi, A., Doyle, B., Gini, F., Guinamard, K., Murakonda, S. K., Liddell, J., Passerat-Palmbach, J., & Weller, S. (2024). Starlit: Privacy-preserving federated learning to enhance financial fraud detection (arXiv:2401.10765). arXiv. https://doi.org/10.48550/arXiv.2401.10765

[21]. Haseeb, A., Ekerete, I., & Moore, S. (2024). A privacy-preserving federated learning framework for financial crime. In International Conference on Ubiquitous Computing and Ambient Intelligence (pp. 743–754). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-77232-0_68

[22]. He, P., Lin, C., & Montoya, I. (2024). DPFedBank: Crafting a privacy-preserving federated learning framework for financial institutions with policy pillars (arXiv:2410.13753). arXiv. https://doi.org/10.48550/arXiv.2410.13753

[23]. [23] Salam, M. A., Fouad, K. M., Elbably, D. L., & Elsayed, S. M. (2024). Federated learning model for credit card fraud detection with data balancing techniques. Neural Computing and Applications, 36, 6231–6256. https://doi.org/10.1007/s00521-023-09410-2
[24]. Zhang, S., Tay, J., & Baiz, P. (2024). The effects of data imbalance under a federated learning approach for credit risk forecasting (arXiv:2401.07234). arXiv. https://doi.org/10.48550/arXiv.2401.07234
[25]. Arora, S., Beams, A., Chatzigiannis, P., Meiser, S., Patel, K., Raghuraman, S., Schoppmann, P., & Zamani, M. (2023). Privacy-preserving financial anomaly detection via federated learning & multi-party computation (arXiv:2310.04546). arXiv. https://doi.org/10.48550/arXiv.2310.04546
[26]. Kadhe, S. R., Ludwig, H., Baracaldo, N., King, A., Zhou, Y., Houck, K., & Soceanu, O. (2023). Privacy-preserving federated learning over vertically and horizontally partitioned data for financial anomaly detection (arXiv:2310.19304). arXiv. https://doi.org/10.48550/arXiv.2310.19304
[27]. Zhang, H., Hong, J., Dong, F., Drew, S., Xue, L., & Zhou, J. (2023). A privacy-preserving hybrid federated learning framework for financial crime detection (arXiv:2302.03654). arXiv. https://doi.org/10.48550/arXiv.2302.03654
[28]. Kanamori, S., Abe, T., Ito, T., Emura, K., Wang, L., Yamamoto, S., & Moriai, S. (2022). Privacy-preserving federated learning for detecting fraudulent financial transactions in Japanese banks. Journal of Information Processing, 30, 789–795. https://doi.org/10.2197/ipsjjip.30.789
[29]. Byrd, D., & Polychroniadou, A. (2020). Differentially private secure multi-party computation for federated learning in financial applications. Proceedings of the First ACM International Conference on AI in Finance, 1–9. https://doi.org/10.1145/3383455.3422562
[30]. Li, Z., Sharma, V., & Mohanty, S. P. (2020). Preserving data privacy via federated learning: Challenges and solutions. IEEE Consumer Electronics Magazine, 9(3), 8–16. https://doi.org/10.1109/MCE.2019.2959108
[31]. Moon, S., & Lee, W. H. (2023). Privacy-preserving federated learning in healthcare. 2023 International Conference on Electronics, Information, and Communication (ICEIC), 1–4. https://doi.org/10.1109/ICEIC57457.2023.10049966
[32]. Ali, M., Naeem, F., Tariq, M., & Kaddoum, G. (2023). Federated learning for privacy preservation in smart healthcare systems: A comprehensive survey. IEEE Journal of Biomedical and Health Informatics, 27(2), 778–789. https://doi.org/10.1109/JBHI.2022.3181823
[33]. Aouedi, O., Sacco, A., Piamrat, K., & Marchetto, G. (2023). Handling privacy-sensitive medical data with federated learning: Challenges and future directions. IEEE Journal of Biomedical and Health Informatics, 27(2), 790–803. https://doi.org/10.1109/JBHI.2022.3185673
[34]. Firdaus, M., & Rhee, K.-H. (2023). Towards trustworthy collaborative healthcare data sharing. IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 4059–4064. https://doi.org/10.1109/BIBM58861.2023.10385319
[35]. Islam, T. U., Ghasemi, R., & Mohammed, N. (2022). Privacy-preserving federated learning model for healthcare data. 2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC), 0281–0287. https://doi.org/10.1109/CCWC54503.2022.9720752
[36]. Narmadha, K., & Varalakshmi, P. (2022). Federated learning in healthcare: A privacy-preserving approach. In Challenges of Trustable AI and Added-Value on Health (pp. 194–198). IOS Press. https://doi.org/10.3233/SHTI220432
[37]. Chowdhury, A., Kassem, H., Padoy, N., Umeton, R., & Karargyris, A. (2021). A review of medical federated learning: Applications in oncology and cancer research. In International MICCAI Brainlesion Workshop (pp. 3–24). Springer International Publishing. https://doi.org/10.1007/978-3-031-08999-2_1
[38]. Pfitzner, B., Steckhan, N., & Arnrich, B. (2021). Federated learning in a medical context: A systematic literature review. ACM Transactions on Internet Technology (TOIT), 21(2), 1–31. https://doi.org/10.1145/3412357