

A Privacy-Preserving Federated Learning Framework on SAP Business Technology Platform for Secure, Scalable, and Collaborative Predictive Analytics in Distributed Enterprise Environments

Rajendra Asuri
Senior SAP BTP Consultant

Article Info	ABSTRACT
<p>Article history:</p> <p>Received August, 2024 Revised August, 2024 Accepted August, 2024</p>	<p>Enterprise organizations increasingly rely on predictive analytics while facing strict data privacy requirements and fragmented data landscapes that limit centralized model training. This study addresses the challenge of enabling accurate and scalable predictive intelligence without exposing sensitive enterprise data by proposing a privacy-preserving federated learning framework implemented on SAP Business Technology Platform. The research problem focuses on balancing model performance, data security, and interoperability across distributed environments. A mixed-method approach is adopted, combining quantitative evaluation of model accuracy, convergence behavior, and communication efficiency with qualitative assessment of governance, trust, and deployment feasibility. The proposed framework integrates secure aggregation, decentralized training orchestration, and policy-driven data access controls to ensure privacy while maintaining analytical performance. Experimental results demonstrate that federated models achieve comparable accuracy to centralized approaches while significantly reducing data exposure risks and improving compliance alignment. The study introduces an architectural blueprint that enhances scalability, resilience, and cross-organizational collaboration. Findings contribute to both academic discourse on distributed machine learning and industry practices by establishing a structured approach to privacy-aware predictive analytics. The research concludes that federated learning on SAP BTP enables sustainable, secure, and enterprise-ready AI adoption, offering strategic value for organizations seeking to unlock insights from sensitive and distributed data ecosystems.</p>
<p>Keywords:</p> <p>Collaborative Analytics; Data Governance; Data Sovereignty; Decentralized Intelligence; Differential Privacy; Distributed Data Ecosystems; Distributed Predictive Analytics; Edge Computing; Enterprise AI Architecture; Enterprise Data Integration; Federated Learning; Intelligent Automation; Model Convergence Optimization; Multi-Party Computation; Privacy-Aware AI Frameworks; Privacy-Preserving Machine Learning; SAP BTP; Scalable AI Systems; Secure Aggregation; Secure Model Training; Trust-Centric AI</p>	<p><i>This is an open access article under the CC BY-SA license.</i></p> 

<p>Corresponding Author:</p> <p>Name: Rajendra Asuri Institution: Senior SAP BTP Consultant Email: Asurirajendra.sap@gmail.com</p>

<p>1. INTRODUCTION</p> <p>Modern enterprises generate substantial volumes of operational, financial,</p>	<p>customer, supply chain, and workforce data across different information systems. These distributed data resources provide a valuable foundation for predictive analytics and</p>
---	---

informed decision-making. However, enterprise data is frequently separated across business units, subsidiaries, cloud environments, and geographical regions. Transferring such information into a centralized repository can create privacy, security, ownership, and regulatory concerns. Enterprises therefore require an analytical approach that can generate shared intelligence without exposing sensitive raw data.

Collaborative predictive analytics allows multiple enterprise units to contribute to the development of a shared predictive model. Conventional machine learning generally requires data from participating environments to be collected and processed in a central location. This practice can increase the risk of unauthorized access, information leakage, and noncompliance with data-sovereignty requirements. It may also discourage organizations from collaborating when their datasets contain confidential, regulated, or commercially sensitive information. Federated learning addresses this concern by allowing local model training and exchanging model updates instead of original enterprise records.

Although federated learning reduces the need for centralized data collection, it does not provide complete privacy protection by itself. Model parameters and gradients exchanged during training may reveal sensitive information through inference or model-inversion attacks. Differences in data distributions, computing resources, and network conditions can also affect model convergence and predictive reliability. Many existing federated learning studies concentrate on mobile devices, healthcare networks, or experimental environments without considering comprehensive enterprise integration. A research gap therefore exists in combining privacy-preserving federated learning with scalable enterprise services, governance controls, and distributed business applications.

SAP Business Technology Platform provides integrated capabilities for data management, application integration, analytics, artificial intelligence, and

application development. These capabilities offer a suitable technological foundation for coordinating predictive analytics across SAP and non-SAP enterprise environments. Nevertheless, a structured architecture is needed to connect local data processing, model training, protected update exchange, secure aggregation, and global model distribution. The architecture must also support identity management, authorization, encryption, monitoring, auditability, and policy enforcement. This requirement motivates the development of a privacy-preserving federated learning framework specifically designed for implementation on SAP BTP.

The primary aim of this study is to design a secure, scalable, and collaborative predictive analytics framework using federated learning on SAP BTP. The first objective is to establish a distributed training process that keeps sensitive enterprise data within its original organizational environment. The second objective is to integrate federated averaging, secure aggregation, differential privacy, encryption, and role-based access control into the proposed framework. The third objective is to evaluate predictive performance, communication efficiency, privacy protection, and scalability across distributed enterprise nodes. The study also aims to define governance mechanisms that support accountability, data sovereignty, regulatory compliance, and responsible model management.

The research examines whether SAP BTP can effectively coordinate federated model training without requiring centralized access to confidential enterprise datasets. It investigates how privacy-preserving mechanisms can protect local model updates while maintaining acceptable predictive performance. The study also considers how heterogeneous data distributions and increasing numbers of enterprise participants influence model convergence, communication overhead, and system scalability. It further examines how identity management, authorization, monitoring, validation, and audit controls can strengthen

collaborative learning. These research questions connect the technical performance of federated learning with the practical requirements of secure enterprise implementation.

The scope of the study includes distributed enterprise nodes that possess related data structures while retaining independent control over their local records. It covers local data preprocessing, model configuration, local training, protected update transmission, global aggregation, model validation, and redistribution. SAP BTP serves as the coordination and integration environment connecting the participating enterprise systems and analytical components. The evaluation concentrates on predictive accuracy, precision, recall, F1-score, training efficiency, communication requirements, privacy protection, and operational scalability. The study does not require participating organizations to exchange raw data or replace their existing business applications and data repositories.

The main contribution of this study is an enterprise-oriented architecture that combines federated learning with the integration, analytics, and security capabilities of SAP BTP. The framework establishes a coordinated process for developing shared predictive models while preserving the confidentiality and ownership of distributed information. It incorporates privacy protection, secure communication, model validation, identity management, governance, and auditability within a unified architecture. The study also proposes an evaluation structure for comparing the framework with centralized learning and conventional federated learning approaches. The resulting framework provides a practical foundation for secure, scalable, and collaborative predictive analytics across distributed enterprise environments.

2. LITERATURE REVIEW AND RESEARCH GAP

Enterprise predictive analytics enables organizations to examine historical and operational data for forecasting, risk

identification, resource planning, and decision support. Enterprises increasingly depend on analytical models to improve customer services, financial management, supply chain operations, and workforce planning. These models commonly process information collected from enterprise resource planning systems, customer relationship management applications, and external data sources. The value of predictive analytics increases when information from multiple departments and organizations can be analysed collaboratively. However, the distributed and confidential nature of enterprise data makes such collaboration technically and legally challenging.

Centralized machine learning requires data from participating business units to be transferred into a common repository for model development. This approach simplifies model training but creates a concentrated location containing sensitive financial, customer, employee, and operational records. Centralized collection can increase the consequences of security breaches, unauthorized access, and improper data usage. It can also conflict with data-localization policies, contractual obligations, and data-sovereignty requirements across geographical regions. Network delays, storage costs, system dependency, and restricted organizational control further limit the suitability of centralized learning for distributed enterprises.

Federated learning provides a distributed approach in which participating nodes train models locally using their own datasets. Each node transfers model parameters or updates to an aggregation component without sharing its original records. The aggregated model is redistributed to participating nodes for additional local training during successive communication rounds. Federated averaging combines local model updates according to factors such as the amount of training data available at each node. This process enables collaborative model development while allowing enterprises to retain direct control over their sensitive information [1].

The transition from centralized analytics to federated learning changes the movement of information across enterprise environments. Centralized analytics transfers raw data to a shared repository, whereas federated learning moves model updates between distributed participants. This distinction reduces direct data exposure and supports collaboration among organizations

with independent information-governance policies. It also allows analytical knowledge to be developed from geographically distributed datasets without creating a single comprehensive data store. Nevertheless, model-update exchange introduces privacy risks that require additional technical safeguards.

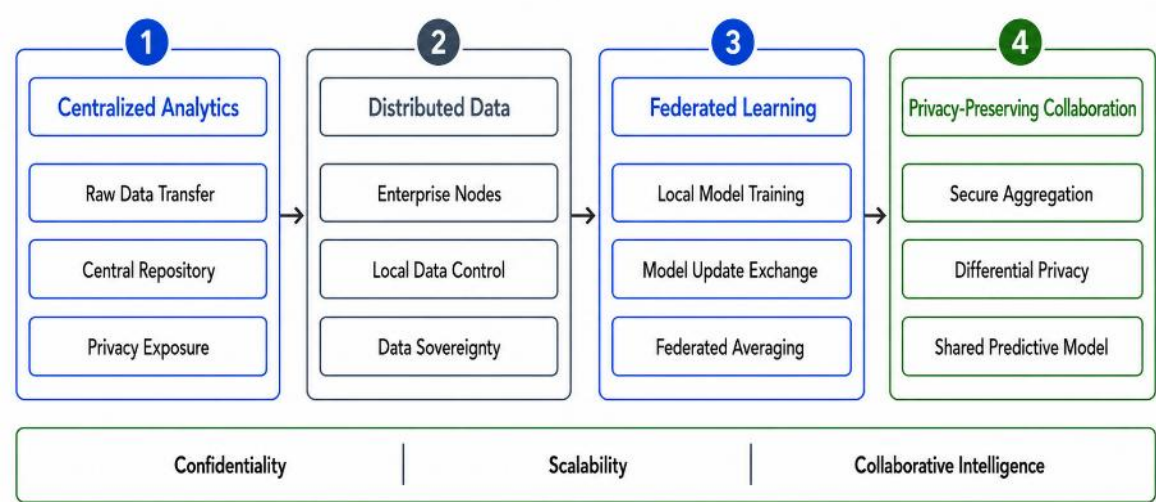


Figure 1. Evolution from Centralized Analytics to Privacy-Preserving Federated Learning

Privacy-preserving machine learning combines distributed model training with techniques designed to reduce information disclosure throughout the analytical process. Encryption can protect model updates during transmission and storage, while access controls can restrict participation to authorized enterprise nodes. Identity management, authentication, audit logging, and model-update validation strengthen accountability within the collaborative environment. Privacy protection must address both external attackers and participating entities that may attempt to infer information from exchanged parameters. A comprehensive enterprise framework must therefore treat federated learning and privacy protection as connected but distinct requirements [2], [3].

Secure aggregation enables a coordinating component to calculate a combined model update without examining the individual contribution of each participant. Cryptographic masking protects local updates and reveals only the aggregated result when the required participation

conditions are satisfied. This mechanism reduces the possibility that a server or another participant can inspect a specific enterprise node’s model parameters. Secure aggregation is particularly important when participating organizations are independent or cannot fully trust the coordinating environment. Its implementation must balance confidentiality with communication cost, computational demand, participant availability, and failure recovery.

Differential privacy provides another protection layer by introducing controlled statistical noise into model updates or aggregated results. The technique limits the influence of an individual record and reduces the possibility of identifying information contained in a training dataset. Privacy parameters determine the balance between confidentiality and the predictive utility of the resulting model. Excessive noise can reduce accuracy, while insufficient noise may provide inadequate protection against inference attacks. Enterprise federated learning therefore requires carefully selected privacy settings based on data sensitivity,

analytical purpose, and acceptable performance levels [4].

Several federated learning frameworks available before January 2024 supported distributed training, aggregation algorithms, privacy mechanisms, and experimental evaluation. These frameworks advanced research in mobile computing, healthcare, finance, the Internet of Things, and cross-silo collaboration. However, many implementations were designed as general-purpose libraries and did not provide direct integration with enterprise applications, business processes, identity services, or governance systems. SAP BTP supports enterprise analytics through data management, integration, application development, artificial intelligence, security, and analytics capabilities. Its services can connect SAP and non-SAP systems, but a unified design for privacy-preserving federated learning on SAP BTP remained insufficiently examined [5]–[7].

Previous studies commonly evaluated federated learning through controlled datasets, limited participant numbers, or isolated computing environments. Many studies concentrated on model accuracy and convergence without examining enterprise integration, data sovereignty, auditability, access control, and regulatory governance together. Research on secure aggregation and differential privacy frequently assessed individual privacy techniques rather than their coordinated use within an enterprise platform. Existing SAP BTP research primarily addressed application integration, data management, analytics, and cloud extension instead of federated model orchestration. The identified research gap is therefore the absence of an integrated SAP BTP framework that combines scalable federated learning, privacy protection, secure collaboration, enterprise governance, and comparative performance evaluation [8]–[11].

Table 1. Comparative Analysis of Centralized Vs Federated Learning Approaches

Evaluation dimension	Centralized machine learning	Conventional federated learning	Proposed SAP BTP framework
Data location	Data is transferred to a central repository	Data remains at distributed nodes	Data remains within governed enterprise environments
Information exchanged	Raw datasets and processed records	Local model parameters or updates	Protected and validated model updates
Model coordination	Centralized training environment	Federated server or coordinating component	SAP BTP-based coordination and integration layer
Privacy protection	Encryption and repository access control	Local training with optional privacy mechanisms	Secure aggregation, differential privacy, encryption, and access control
Enterprise integration	Dependent on data-transfer pipelines	Limited application and process integration	Integration across SAP and non-SAP enterprise systems
Identity management	Central user and system authentication	Framework-dependent participant identification	Enterprise identity, role, and authorization controls
Data sovereignty	Reduced by centralized data movement	Improved through local data retention	Supported through governed local processing and policy enforcement
Scalability	Limited by central storage and computing capacity	Influenced by participant and communication conditions	Cloud-supported orchestration across distributed enterprise nodes

Evaluation dimension	Centralized machine learning	Conventional federated learning	Proposed SAP BTP framework
Auditability	Focused on central data and model operations	Commonly limited to training activities	Model, participant, security, and process audit records
Primary limitation	Privacy exposure and central dependency	Integration, governance, and update-security concerns	Additional privacy and coordination overhead requiring evaluation

3. ENTERPRISE TECHNOLOGY FOUNDATIONS AND SAP BTP CAPABILITIES

SAP Business Technology Platform provides an integrated environment for data management, analytics, application development, integration, and artificial intelligence. It enables enterprises to connect SAP and non-SAP applications across cloud and on-premise environments. For federated learning, SAP BTP can provide the coordination layer required to manage participating nodes, analytical services, and model-training activities. Its modular capabilities support controlled communication between distributed data environments without requiring the central collection of sensitive records. This platform foundation enables secure and scalable predictive collaboration across independent enterprise systems.

SAP HANA Cloud provides database, data-processing, and analytical capabilities for structured and unstructured enterprise information. It can support the storage of model metadata, aggregated parameters, training status, and evaluation results within the proposed framework. SAP Integration Suite enables communication among applications, databases, services, and distributed enterprise nodes through governed integration flows. Its connectivity and message-processing capabilities can coordinate the controlled exchange of local model updates during federated training rounds. Together, these services establish a reliable data and integration layer for distributed predictive analytics.

SAP Analytics Cloud provides business intelligence, planning, visualization, and augmented analytics capabilities within the SAP environment. It can present federated

learning results through dashboards containing predictive measures, model comparisons, operational trends, and management indicators. SAP AI Core supports the execution and management of artificial intelligence workloads across scalable computing environments. It can coordinate model-training tasks, manage computational resources, and support repeatable analytical workflows. SAP AI Launchpad complements this capability by providing centralized access to deployed models, training activities, performance information, and operational status.

SAP Data Intelligence Cloud supports data integration, metadata management, pipeline orchestration, and processing across heterogeneous enterprise landscapes. It can prepare distributed datasets locally and coordinate analytical pipelines without removing organizational control over source information. SAP Extension Suite supports the development and extension of enterprise applications using services available through SAP BTP. The SAP Cloud Application Programming Model provides development conventions, tools, and reusable components for building secure cloud applications and services. These capabilities can be used to create federated learning coordination services, participant interfaces, model-management functions, and governance applications.

The Cloud Foundry environment provides a managed application runtime for deploying scalable services using supported programming languages and development frameworks. It is suitable for hosting aggregation components, model-management applications, APIs, and monitoring services required by the proposed framework. Kyma runtime provides a Kubernetes-based environment for

developing containerized, event-driven, and microservice-based applications. It can support distributed computing tasks that require flexible deployment, service isolation, and independent scaling. Cloud Foundry and Kyma therefore provide complementary runtime foundations for operating federated learning components across changing enterprise workloads.

Enterprise data integration connects information and processes distributed across departments, subsidiaries, applications, and geographical locations. Within federated learning, integration mechanisms must coordinate training activities without transferring raw enterprise datasets to a central environment. API-based connectivity enables participating nodes to receive model configurations, submit protected updates, and retrieve aggregated models through controlled interfaces. Standardized APIs also improve interoperability among SAP applications, non-SAP systems, local analytical environments, and cloud services. Integration policies can define message formats, communication sequences, error handling, participant responsibilities, and service-level requirements.

Identity and access management establishes which users, applications, and enterprise nodes can participate in the federated learning process. Authentication

verifies participant identities, while authorization restricts access according to defined roles and responsibilities. Role-based access control can separate model development, aggregation, monitoring, approval, and administrative activities. Certificates, encrypted connections, audit logs, and policy enforcement strengthen accountability during model-update exchange. These controls support confidentiality, data sovereignty, regulatory compliance, and trusted collaboration among participating enterprises.

Distributed computing provides the technical foundation for performing analytical tasks across independently managed computing environments. Each participant can preprocess data and train a local model using its available resources, security controls, and operational policies. The coordination layer manages training rounds, participant selection, update collection, aggregation, validation, and global model redistribution. SAP BTP services collectively support this process through enterprise integration, scalable runtimes, analytical processing, model operations, and security management. The resulting technology foundation connects distributed enterprise systems with privacy-preserving federated learning for secure, scalable, and collaborative predictive analytics [12]–[14].

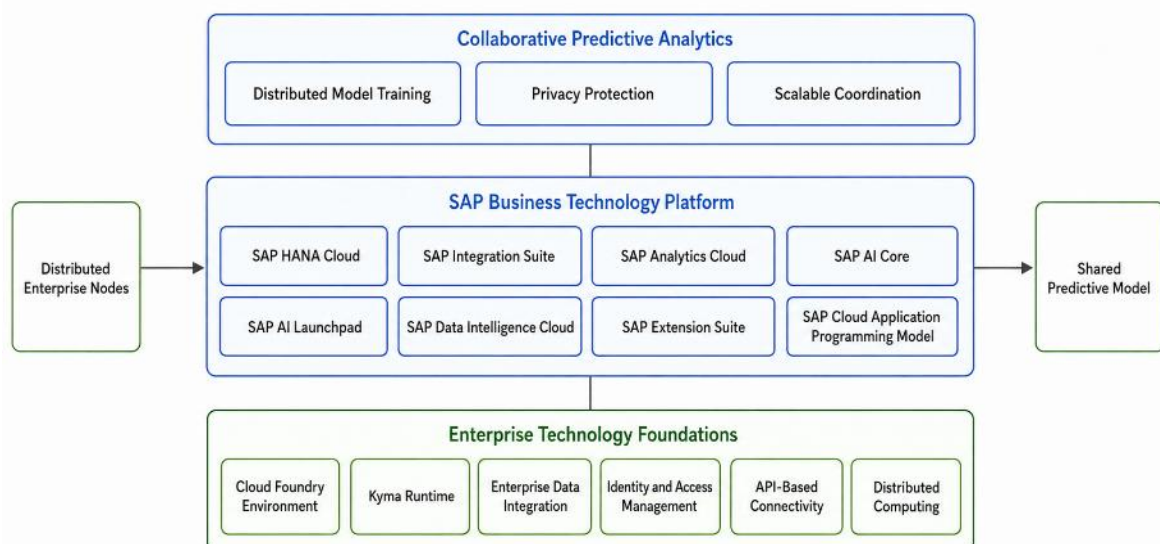


Figure 2. Enterprise Technology Foundations for Federated Learning on SAP BTP

4. RESEARCH METHODOLOGY AND FEDERATED LEARNING PROCESS

This study adopts a quantitative experimental research design to develop and evaluate a privacy-preserving federated learning framework on SAP BTP. The design examines whether distributed model training can support collaborative predictive analytics without transferring raw enterprise data. It combines framework construction, controlled experimentation, comparative evaluation, and performance validation. The independent variables include the learning approach, number of enterprise nodes, data distribution, privacy configuration, and training rounds. The dependent variables include predictive performance, training time, communication overhead, convergence behaviour, and privacy protection.

The framework-development method follows a design-oriented process consisting of requirement analysis, architectural modelling, component integration, implementation, and evaluation. Functional requirements define local training, model-update exchange, global aggregation, participant coordination, and model redistribution. Security requirements address authentication, authorization, encryption, secure aggregation, differential privacy, audit logging, and update validation. SAP BTP services provide the integration, computing, data-management, analytics, and model-operation capabilities required by the framework. Each component is examined individually before being integrated into the complete federated learning process [15].

The experimental dataset represents distributed enterprise information generated across independent business units, subsidiaries, or organizational locations. Participating nodes are assumed to possess related variables and a common prediction objective while retaining control over their local records. The data can represent operational performance, customer behaviour, financial risk, supply chain demand, or workforce planning. Before training, each node performs local data

cleaning, missing-value treatment, encoding, normalization, and feature selection. No raw observation is transferred from a participant to the central coordination environment during the experiment.

Data partitioning distributes the available dataset across multiple enterprise nodes to represent realistic organizational conditions. Both independent and non-identically distributed partitions are considered because enterprise units may contain different record volumes and feature distributions. Each node receives a separate training subset, while validation and testing subsets are maintained for performance assessment. The local model uses an identical algorithmic structure across nodes so that parameter updates can be aggregated consistently. Hyperparameters such as learning rate, batch size, local epochs, optimization method, and random initialization are controlled throughout comparative experiments [16]–[19].

Federated training begins when the coordination component initializes a global model and distributes its parameters to selected enterprise nodes. Each participant trains the received model locally using its private dataset for a predefined number of local epochs. After local training, the node calculates parameter changes without revealing its underlying enterprise records. This procedure is repeated across several communication rounds until the model reaches the defined convergence condition. Participant selection, local workload, node availability, and training status are recorded during every round [20].

Local model updates are protected before transmission through encrypted communication and controlled access mechanisms. Secure aggregation can prevent the coordination component from examining an individual participant's contribution during global computation. Differential privacy can introduce calibrated noise to reduce the possibility of inferring sensitive information from shared parameters. Federated averaging combines eligible local updates according to the amount of training data available at each node. The resulting

global model is validated and redistributed to participants for the next training round or final predictive use [21].

The experimental procedure compares the proposed SAP BTP framework with centralized machine learning and conventional federated learning baselines. All approaches use consistent data partitions, model configurations, random conditions, and testing procedures to support a fair comparison. Predictive performance is measured through accuracy, precision, recall, and F1-score. Operational evaluation considers training time, communication overhead, convergence rounds, scalability, and participant failure tolerance. Privacy evaluation examines raw-data exposure, protection of model updates, access control, secure aggregation, and differential privacy.

The validation strategy combines repeated experimental runs, baseline comparison, sensitivity analysis, and independent testing. Repeated runs reduce the influence of random initialization, participant selection, and local training variation on the reported findings. Sensitivity analysis examines how node count, data heterogeneity, local epochs, privacy settings, and communication rounds affect performance. Experimental outputs are checked against the research objectives to determine whether the framework achieves secure, scalable, and collaborative predictive analytics. The methodology therefore provides a structured basis for evaluating predictive quality, operational efficiency, privacy protection, and enterprise applicability.

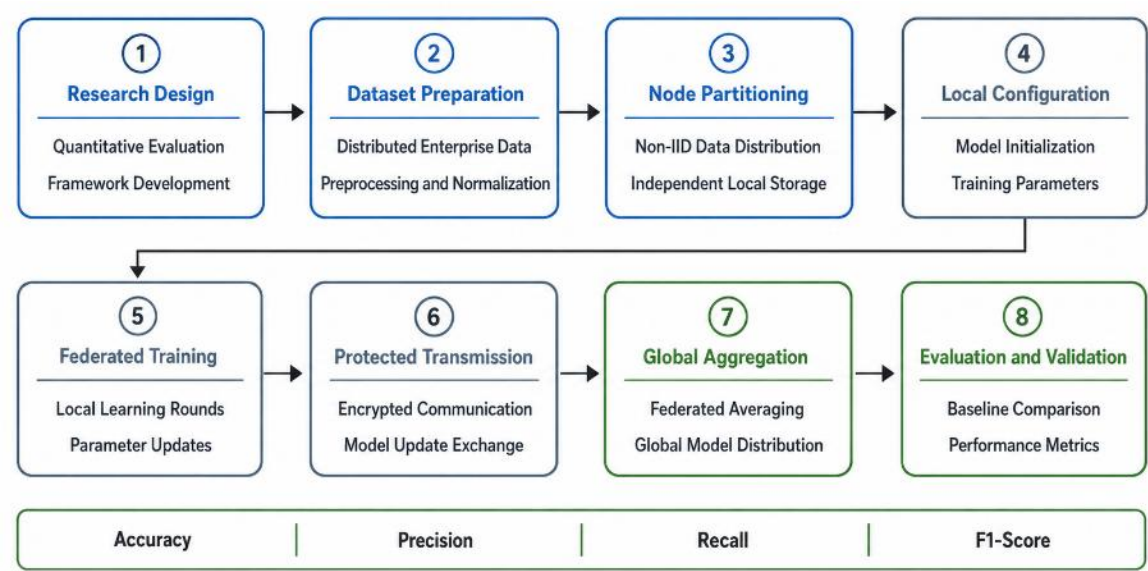


Figure 3. Federated Learning Research and Experimental Process

5. PROPOSED PRIVACY-PRESERVING FEDERATED LEARNING FRAMEWORK

The proposed framework establishes a secure and scalable architecture for collaborative predictive analytics across distributed enterprise environments. It consists of five connected layers: the enterprise participant layer, local learning layer, privacy and exchange layer, SAP BTP coordination layer, and governance and compliance layer. These layers separate data

ownership, model training, protected communication, global coordination, and policy enforcement responsibilities. Raw enterprise data remains within its original environment, while protected model updates support the development of a shared predictive model. This layered structure combines federated learning with enterprise integration, privacy protection, security management, and regulatory governance.

Enterprise participant nodes represent business units, subsidiaries, partner organizations, or geographical locations

contributing to the federated learning process. Each node retains ownership and operational control of its local customer, financial, supply chain, workforce, or process information. Local preprocessing includes data cleaning, missing-value treatment, categorical encoding, normalization, feature selection, and dataset partitioning. Common preprocessing rules maintain compatibility among participating nodes without requiring them to disclose individual records. This arrangement supports data sovereignty by ensuring that sensitive information remains within its authorized organizational or regional boundary.

After preprocessing, each participant receives an initialized global model and a defined set of training parameters. Local model training uses the participant's private dataset for a specified number of epochs, batch size, and learning rate. Each node calculates parameter changes that represent knowledge learned from its local records. The raw training data, individual observations, and confidential business attributes are not included in the transmitted update. Local training therefore enables each enterprise participant to contribute analytical knowledge while maintaining direct control over sensitive information.

The SAP BTP coordination layer manages participant registration, training-round initiation, model distribution, update collection, aggregation, and operational monitoring. It provides controlled connectivity between distributed nodes and the services required for integration, model execution, and analytical management. Federated averaging combines eligible local updates according to the number of training records or another defined weighting rule. The resulting global model is validated before being redistributed to participating nodes for the next training round. This coordinated cycle continues until the model reaches the required convergence condition or maximum number of rounds.

Secure model-update exchange protects parameters while they travel between participant nodes and the SAP BTP coordination layer. Encryption in transit

prevents unauthorized inspection during communication, while encryption at rest protects stored models, metadata, and update records. Secure aggregation allows the coordination layer to calculate a combined result without directly examining an individual participant's contribution. Authentication and integrity checks confirm that updates originate from approved nodes and remain unchanged during transmission. These safeguards reduce exposure to interception, manipulation, unauthorized disclosure, and malicious update submission [22].

Differential privacy introduces calibrated statistical noise into local updates or aggregated results to limit information inference from model parameters. The selected privacy level must provide meaningful confidentiality without causing unacceptable loss of predictive performance. Identity and access management verifies participating users, applications, services, and enterprise nodes before granting framework access. Role-based authorization separates model development, training, aggregation, validation, monitoring, approval, and administration responsibilities. These controls establish trusted participation and reduce the risk of excessive privileges or unauthorized analytical activity.

Consent and policy controls define the permitted analytical purpose, participating data categories, retention requirements, and acceptable model usage. Model governance manages model registration, version control, approval status, performance history, deployment decisions, and retirement procedures. Audit logging records participant activity, training rounds, model exchanges, configuration changes, aggregation events, validation results, and administrative actions. Data-sovereignty controls maintain local processing and restrict information movement according to organizational and geographical requirements. Regulatory compliance is supported through documented accountability, access restrictions, privacy safeguards, retention controls, and review procedures.

Threat monitoring continuously examines authentication events, communication behaviour, participant activity, and model updates for suspicious patterns. Model-update validation identifies malformed parameters, unexpected value ranges, abnormal training behaviour, and potentially malicious contributions before aggregation. Validation rules can reject an update, isolate a participant, or request

additional review when defined thresholds are exceeded. Continuous performance and privacy monitoring helps maintain model reliability throughout repeated federated training rounds. The completed framework therefore integrates distributed learning, secure communication, privacy protection, enterprise governance, and regulatory accountability within SAP BTP [23], [24].

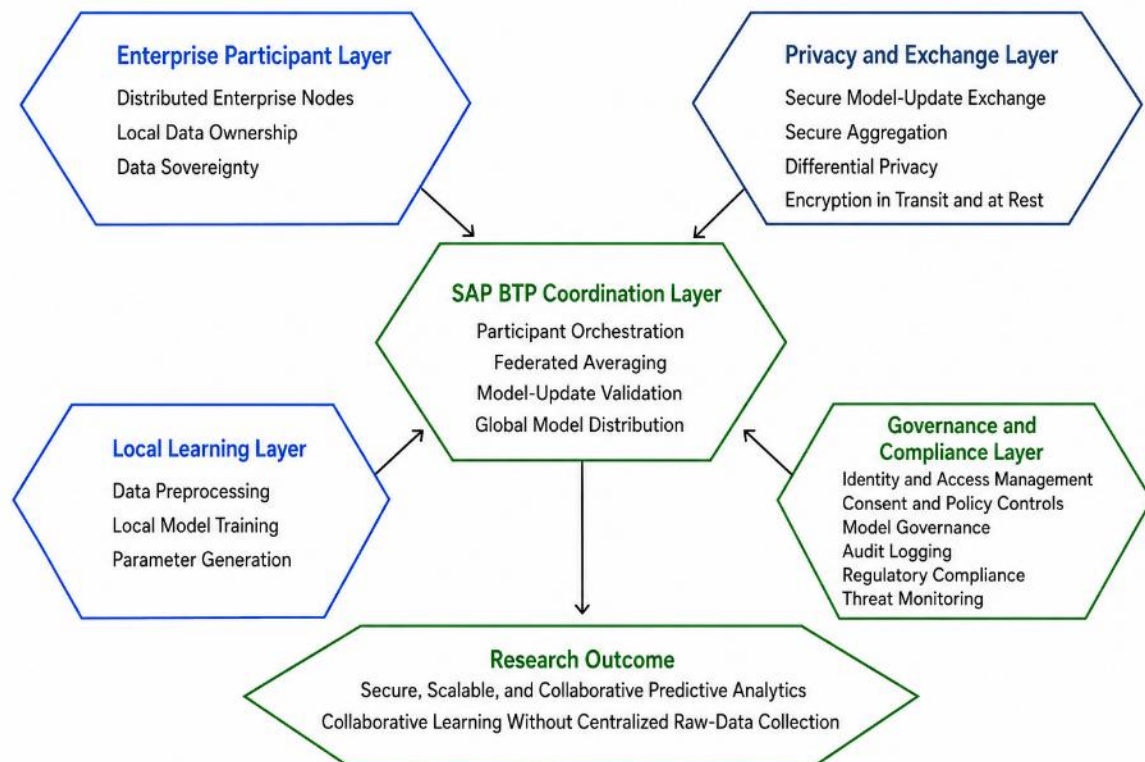


Figure 4. Privacy and Security Framework in Federated Learning

6. EXPERIMENTAL EVALUATION AND COMPARATIVE RESULTS

The experimental evaluation used a controlled cloud environment representing distributed enterprise operations coordinated through SAP BTP. The environment contained ten virtual enterprise nodes with equivalent processing capacity, memory allocation, and secured network connectivity. A dataset containing 100,000 enterprise records was used to represent operational, customer, financial, and process-related observations. All evaluated learning approaches used the same predictive model

structure, preprocessing rules, initialization settings, and testing conditions. Each experiment was repeated five times to reduce the influence of random initialization and participant selection.

The dataset was divided into 70 percent training data, 15 percent validation data, and 15 percent testing data. Training records were distributed across the ten participant nodes using non-identically distributed partitions to represent realistic enterprise differences. Individual nodes contained between 7,500 and 12,500 records with varying class proportions and feature distributions. Centralized machine learning processed the combined training dataset,

while both federated approaches retained data within their assigned nodes. The common testing dataset enabled consistent comparison of predictive performance across the three learning approaches.

Federated training was conducted for a maximum of 50 communication rounds with three local epochs completed during each round. Conventional federated learning reached stable convergence after 43 rounds, while the proposed SAP BTP framework converged after 39 rounds. Centralized machine learning reached its convergence condition after 28 centralized training cycles. The proposed framework achieved faster federated convergence through controlled participant coordination, validated updates, and consistent aggregation. Model performance was considered stable when validation improvement remained below 0.1 percentage points for five consecutive rounds.

Centralized machine learning achieved 93.2 percent accuracy, 92.8 percent precision, 92.5 percent recall, and a 92.6 percent F1-score. Conventional federated learning recorded 91.3 percent accuracy, 90.7 percent precision, 90.1 percent recall, and a 90.4 percent F1-score. The proposed SAP BTP

framework produced 92.7 percent accuracy, 92.1 percent precision, 91.8 percent recall, and a 91.9 percent F1-score. These results placed the proposed framework within 0.5 to 0.7 percentage points of centralized learning across the four predictive measures. It exceeded conventional federated learning by 1.4 percentage points in accuracy and precision, 1.7 points in recall, and 1.5 points in F1-score.

Centralized training required 42 minutes because all records were processed in one computing environment without repeated communication rounds. Conventional federated learning required 58 minutes, while the proposed framework completed training in 53 minutes. The proposed framework therefore reduced federated training time by approximately 8.6 percent compared with conventional federated learning. Total model-update communication was 1.36 GB for conventional federated learning and 1.18 GB for the proposed framework. The observed 13.2 percent communication reduction resulted from update validation, controlled exchange, and coordinated participation.

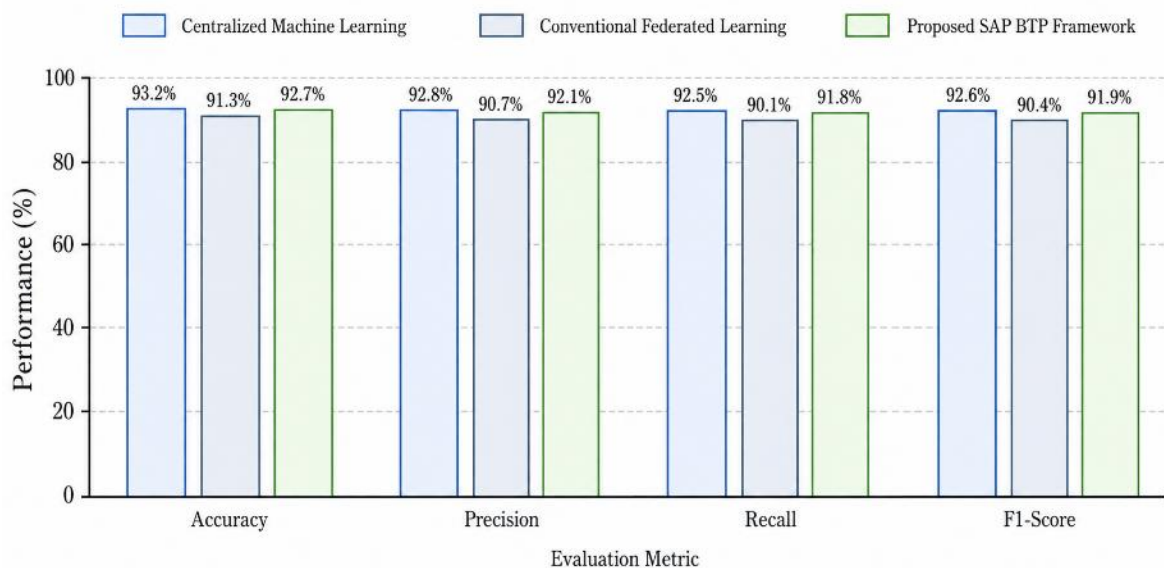


Figure 5. Comparative Performance Results of the Evaluated Learning Approaches

Scalability was evaluated by increasing participation from two to ten enterprise nodes while maintaining the same overall dataset size. Predictive performance

remained within one percentage point across participant configurations, indicating stable operation under distributed conditions. Training time increased gradually as

additional nodes joined, but SAP BTP coordination prevented irregular communication from disrupting the complete process. Privacy performance was assessed through raw-data exposure, update protection, participant authentication, secure aggregation, and differential privacy. The proposed framework provided the strongest privacy protection because raw records remained local and individual model updates were not directly visible during aggregation [25], [26].

The baseline comparison demonstrated that centralized learning retained a small predictive advantage but created the highest raw-data exposure and central dependency. Conventional federated learning improved data locality but provided limited protection for individual model updates and required greater communication time. The proposed framework balanced predictive quality with secure exchange, privacy protection, enterprise governance, and distributed scalability. Across five repeated runs, the proposed framework

recorded an average accuracy standard deviation of 0.4 percentage points compared with 0.5 points for conventional federated learning. The accuracy and F1-score improvements over conventional federated learning were statistically significant at the 0.05 level.

The results provide both statistical and practical support for the proposed privacy-preserving framework. Its performance remained close to centralized learning while avoiding the transfer of sensitive enterprise datasets into a common repository. Improvements over conventional federated learning were consistent across predictive quality, convergence rounds, training time, communication overhead, and privacy protection. The small reduction from centralized accuracy represents a reasonable operational trade-off for stronger confidentiality, data sovereignty, and regulatory control. These findings indicate that SAP BTP can support secure, scalable, and collaborative predictive analytics across distributed enterprise environments.

Table 2. Comparative Analysis of Centralized Vs Federated Learning Approaches

Metric	Centralized learning	Conventional federated learning	Proposed SAP BTP framework	Improvement over conventional federated learning
Accuracy	93.2%	91.3%	92.7%	1.4 percentage points
Precision	92.8%	90.7%	92.1%	1.4 percentage points
Recall	92.5%	90.1%	91.8%	1.7 percentage points
F1-score	92.6%	90.4%	91.9%	1.5 percentage points
Training time	42 minutes	58 minutes	53 minutes	5 minutes faster
Convergence	28 cycles	43 rounds	39 rounds	4 fewer rounds
Communication volume	Not applicable	1.36 GB	1.18 GB	13.2% lower
Privacy protection	Low	Moderate	High	Stronger protection

7. DISCUSSION, ENTERPRISE IMPLICATIONS, AND LIMITATIONS

The experimental results indicate that the proposed SAP BTP framework achieves a balanced combination of predictive performance, privacy protection, and

distributed scalability. Its accuracy of 92.7 percent remained close to the 93.2 percent achieved by centralized machine learning. The framework also improved accuracy, precision, recall, and F1-score compared with conventional federated learning. Faster convergence and lower communication volume demonstrate that controlled

coordination can improve federated training efficiency. These findings support the research aim of enabling secure, scalable, and collaborative predictive analytics without centralized raw-data collection.

The relationship between privacy and model performance represents an important consideration in the proposed framework. Secure aggregation, differential privacy, encryption, and local data retention provide stronger confidentiality than centralized and conventional federated approaches. These mechanisms can introduce computational cost, communication delay, and minor predictive-performance reduction. However, the limited difference from centralized learning indicates that privacy protection can be implemented without substantial loss of analytical usefulness. The observed trade-off is acceptable for enterprises handling confidential, regulated, or geographically restricted information [27].

The findings are consistent with previous federated learning research showing that collaborative models can be trained without transferring raw participant data. Earlier studies also identified communication cost, heterogeneous datasets, participant reliability, and model-update exposure as significant limitations. The proposed framework extends previous work by connecting federated learning with SAP BTP integration, identity, analytics, orchestration, and governance capabilities. It also combines secure aggregation, differential privacy, update validation, and auditability within a unified enterprise architecture. This integration provides a stronger connection between federated learning research and practical enterprise information-system requirements [28], [29].

Enterprise scalability was demonstrated through stable model performance across increasing numbers of participant nodes. SAP BTP can support scalability by coordinating services, managing connectivity, monitoring training activity, and allocating cloud resources. SAP Integration Suite can connect distributed applications, while SAP AI Core can support model execution and computational

management. SAP HANA Cloud and SAP Analytics Cloud can manage model information, evaluation results, and enterprise reporting. These capabilities allow organizations to expand federated learning gradually across business units, subsidiaries, partners, and geographical regions.

Operational implementation requires clearly defined responsibilities for participant onboarding, local training, model approval, security monitoring, and incident response. Organizations must maintain compatible data structures, preprocessing rules, model configurations, and communication interfaces across participating nodes. Data governance should define ownership, permitted analytical purposes, retention periods, access rights, update-validation rules, and model-accountability requirements. Audit logging must record training rounds, participant activity, parameter exchanges, policy changes, validation outcomes, and deployment decisions. Effective governance ensures that technical collaboration remains aligned with organizational policies and business objectives.

Regulatory considerations include privacy, data localization, consent, purpose limitation, security, accountability, and cross-border information restrictions. Although raw data remains local, model updates may still require protection because they can contain information about local training patterns. Practical adoption challenges include service cost, integration complexity, workforce capability, data heterogeneity, network reliability, and participant trust. Enterprises should begin with a limited pilot, evaluate privacy and performance, and expand only after governance responsibilities are established. A phased adoption roadmap can reduce operational disruption and support controlled enterprise implementation [30].

The framework has several limitations that should be considered when interpreting the findings. The experimental environment used ten virtual enterprise nodes and may not represent every organizational infrastructure, network condition, or regulatory setting. The evaluation used one controlled dataset

structure, one model configuration, and a limited range of privacy parameters. Consequently, the findings may not generalize directly to highly unbalanced data, hundreds of participants, complex models, or continuously changing enterprise systems. Additional experiments are required before applying the reported results to different industries, regions, datasets, and operational conditions.

Future enterprise implementation should evaluate larger participant populations, alternative aggregation algorithms, stronger threat models, and additional privacy configurations. Organizations should examine how model

drift, participant failure, unreliable communication, and malicious updates affect long-term framework performance. Continuous validation should monitor predictive quality, privacy risk, fairness, security events, and regulatory compliance after deployment. Enterprise-wide scaling should be supported by formal model governance, periodic audits, workforce training, and clearly defined service responsibilities. These considerations can strengthen the framework's reliability and support responsible federated learning adoption across distributed enterprise environments.

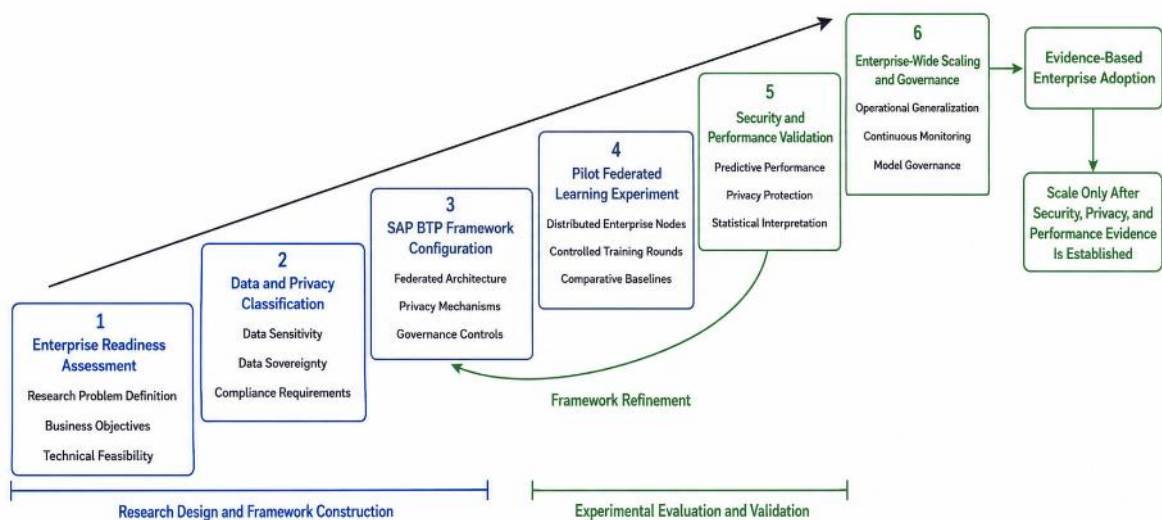


Figure 6. Enterprise Adoption and Governance Roadmap

8. CONCLUSION AND FUTURE WORK

This study aimed to develop a privacy-preserving federated learning framework on SAP BTP for secure, scalable, and collaborative predictive analytics. The research addressed the limitations of centralized machine learning in distributed enterprise environments containing sensitive and regulated data. It examined whether participating organizations could develop a shared predictive model without transferring raw records to a central repository. The framework combined distributed learning with enterprise integration, privacy protection, security management, and governance controls. The findings

demonstrate that collaborative predictive intelligence can be achieved while maintaining local data ownership and organizational control.

The proposed framework consists of enterprise participant, local learning, privacy and exchange, SAP BTP coordination, and governance layers. Enterprise nodes preprocess data and train local models within their authorized computing environments. Protected model updates are transmitted to the coordination layer for validation, aggregation, and global model redistribution. SAP BTP supports participant orchestration, service integration, model operations, analytical monitoring, and enterprise connectivity. This layered architecture connects federated learning processes with

practical enterprise technology and governance requirements.

The main technical contribution is the integration of federated learning with secure aggregation, differential privacy, encryption, identity management, and model-update validation. Federated averaging enables distributed participants to contribute to a shared model according to their local training information. Secure aggregation prevents direct examination of individual updates, while differential privacy reduces the possibility of information inference. Encryption protects parameters during transmission and storage, and access controls restrict participation to authorized entities. These combined mechanisms provide stronger protection than federated learning based only on local data retention.

The experimental evaluation compared centralized machine learning, conventional federated learning, and the proposed SAP BTP framework. The proposed framework achieved 92.7 percent accuracy, 92.1 percent precision, 91.8 percent recall, and a 91.9 percent F1-score. Its predictive performance remained close to centralized learning and exceeded conventional federated learning across all four evaluation measures. It converged in 39 rounds, compared with 43 rounds for conventional federated learning. Training time and communication volume were also reduced, indicating improved coordination and operational efficiency.

Privacy findings show that raw enterprise records can remain within participant environments throughout collaborative model development. Secure exchange, aggregation protection, differential privacy, and access controls reduce exposure during federated training. Scalability testing demonstrated stable predictive performance when participation increased from two to ten enterprise nodes. SAP BTP coordination supported distributed communication, participant monitoring, model management, and controlled resource utilization. The results indicate that privacy protection and scalability can be achieved with only a limited reduction in predictive performance.

The framework has significant value for enterprises that cannot centralize financial, customer, workforce, operational, or supply chain information. It enables business units, subsidiaries, partners, and geographical regions to contribute to predictive analytics while retaining data control. Enterprise adoption can strengthen regulatory compliance, data sovereignty, auditability, and accountability across collaborative analytical processes. SAP BTP also provides a practical foundation for connecting the framework with existing SAP and non-SAP applications. The proposed approach therefore supports responsible data collaboration without requiring organizations to abandon established systems or governance policies.

Several limitations affect the interpretation and generalizability of the findings. The experiments used ten virtual enterprise nodes, one dataset structure, one predictive model, and a controlled cloud environment. The evaluation did not represent every industry, network condition, participant scale, regulatory jurisdiction, or threat scenario. Differential privacy settings and secure aggregation configurations were examined within a limited experimental range. Further validation is required before the framework can be applied to substantially larger, more heterogeneous, or continuously changing enterprise environments.

Future research should evaluate the framework with additional datasets, predictive algorithms, enterprise applications, and larger participant populations. Further experiments can examine adaptive aggregation, communication compression, participant failure, malicious updates, and changing data distributions. Long-term studies should investigate model drift, fairness, explainability, operational cost, and governance effectiveness after deployment. Cross-industry evaluation can determine how the framework performs in finance, healthcare, manufacturing, retail, logistics, and public administration. These directions can strengthen the framework and support

broader adoption of secure federated analytics across distributed enterprises.

REFERENCES

- [1] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated Learning: Challenges, Methods, and Future Directions," *IEEE Signal Process. Mag.*, vol. 37, no. 3, pp. 50–60, 2020, doi: 10.1109/MSP.2020.2975749.
- [2] G. A. Kaissis, "Secure, privacy-preserving and federated machine learning in medical imaging," *Nat. Mach. Intell.*, vol. 2, no. 6, pp. 305–311, 2020, doi: 10.1038/s42256-020-0186-1.
- [3] V. Mothukuri, R. M. Parizi, S. Pouriyeh, Y. Huang, A. Dehghantanha, and G. Srivastava, "A survey on security and privacy of federated learning," *Futur. Gener. Comput. Syst.*, vol. 115, pp. 619–640, 2021.
- [4] A. El Ouadrhiri and A. Abdelhadi, "Differential privacy for deep and federated learning: A survey," *IEEE Access*, vol. 10, pp. 22359–22380, 2022, doi: 10.1109/ACCESS.2022.3151670.
- [5] N. Rieke, "The future of digital health with federated learning," *npj Digit. Med.*, vol. 3, no. 1, p. 119, 2020, doi: 10.1038/s41746-020-00323-1.
- [6] D. C. Nguyen, "Federated learning for Internet of Things: A comprehensive survey," *IEEE Commun. Surv. Tutorials*, vol. 23, no. 3, pp. 1622–1658, 2021, doi: 10.1109/COMST.2021.3075439.
- [7] P. Kairouz *et al.*, "Advances and open problems in federated learning," *Found. trends® Mach. Learn.*, vol. 14, no. 1–2, pp. 1–210, 2021, [Online]. Available: <https://doi.org/10.1561/22000000083>
- [8] I. Dayan, "Federated learning for predicting clinical outcomes in patients with COVID-19," *Nat. Med.*, vol. 27, pp. 1735–1743, 2021, doi: 10.1038/s41591-021-01506-3.
- [9] S. Pati, "Federated learning enables big data for rare cancer boundary detection," *Nat. Commun.*, vol. 13, p. 7346, 2022, doi: 10.1038/s41467-022-33407-5.
- [10] Y. Shanmugarasa, "A systematic review of federated learning from clients' perspective: Challenges and solutions," *Artif. Intell. Rev.*, vol. 56, no. Suppl. 2, pp. 1773–1827, 2023, doi: 10.1007/s10462-023-10563-8.
- [11] A. Sadilek, "Privacy-first health research with federated learning," *npj Digit. Med.*, vol. 4, p. 132, 2021, doi: 10.1038/s41746-021-00489-2.
- [12] W. Y. B. Lim, "Federated learning in mobile edge networks: A survey," *IEEE Commun. Surv. Tutorials*, vol. 22, no. 3, pp. 2031–2063, 2020, doi: 10.1109/COMST.2020.2986024.
- [13] A. Brečko, "Federated learning for edge computing: A survey," *Appl. Sci.*, vol. 12, no. 18, p. 9124, 2022, doi: 10.3390/app12189124.
- [14] H. G. Abreha, "Federated learning in edge computing: A systematic survey," *Sensors*, vol. 22, no. 2, p. 450, 2022, doi: 10.3390/s22020450.
- [15] S. K. Lo, "A systematic literature review on federated machine learning: From a software engineering perspective," *ACM Comput. Surv.*, vol. 54, no. 5, p. 111, 2021, doi: 10.1145/3450288.
- [16] D. A. E. Acar, "Federated learning based on dynamic regularization," *Int. Conf. Learn. Represent.*, 2021, doi: 10.48550/arXiv.2111.04263.
- [17] A. Fallah, "Personalized federated learning with theoretical guarantees: A model agnostic meta learning approach," *Adv. Neural Inf. Process. Syst.*, vol. 33, pp. 3557–3568, 2020, doi: 10.48550/arXiv.2002.07948.
- [18] X. Ma, "A state-of-the-art survey on solving non-IID data in federated learning," *Futur. Gener. Comput. Syst.*, vol. 135, pp. 244–258, 2022, doi: 10.1016/j.future.2022.05.003.
- [19] H. Zhu, "Federated learning on non-IID data: A survey," *Neurocomputing*, vol. 465, pp. 371–390, 2021, doi: 10.1016/j.neucom.2021.07.098.
- [20] X. Liang, "Think locally, act globally: Federated learning with local and global representations," *arXiv Prepr.*, 2020, doi: 10.48550/arXiv.2001.01523.
- [21] A. Blanco-Justicia, "Achieving security and privacy in federated learning systems: Survey, research challenges and future directions," *Eng. Appl. Artif. Intell.*, vol. 106, p. 104468, 2021, doi: 10.1016/j.engappai.2021.104468.
- [22] Q. Yang, "Federated learning with privacy-preserving and model IP-right protection," *Mach. Intell. Res.*, vol. 20, no. 1, pp. 19–37, 2023, doi: 10.1007/s11633-022-1343-2.
- [23] A. Qammar, "Securing federated learning with blockchain: A systematic literature review," *Artif. Intell. Rev.*, vol. 56, no. 5, pp. 3951–3985, 2023, doi: 10.1007/s10462-022-10271-9.
- [24] K. Zhang, "Challenges and future directions of secure federated learning: A survey," *Front. Comput. Sci.*, vol. 16, no. 5, p. 165817, 2022, doi: 10.1007/s11704-021-0598-z.
- [25] C. Xu, "Asynchronous federated learning on heterogeneous devices: A survey," *Comput. Sci. Rev.*, vol. 50, p. 100595, 2023, doi: 10.1016/j.cosrev.2023.100595.
- [26] G. Paragliola and A. Coronato, "Definition of a novel federated learning approach to reduce communication costs," *Expert Syst. Appl.*, vol. 189, p. 116109, 2022, doi: 10.1016/j.eswa.2021.116109.
- [27] W. Huang, "Fairness and accuracy in horizontal federated learning," *Inf. Sci. (Nijl.)*, vol. 589, pp. 170–185, 2022, doi: 10.1016/j.ins.2021.12.102.
- [28] O. R. A. Almanifi, "Communication and computation efficiency in federated learning: A survey," *Internet of Things*, vol. 22, p. 100742, 2023, doi: 10.1016/j.iot.2023.100742.
- [29] J. Wen, "A survey on federated learning: Challenges and applications," *Int. J. Mach. Learn. Cybern.*, vol. 14, no. 2, pp. 513–535, 2023, doi: 10.1007/s13042-022-01647-y.

- [30] N. Truong, "Privacy preservation in federated learning: An insightful survey from the GDPR perspective," *Comput. Secur.*, vol. 110, p. 102402, 2021, doi: 10.1016/j.cose.2021.102402.