Computer Modeling in Engineering & Sciences

Tech Science Press

https://doi.org/10.32604/cmes.2026.083334

ARTICLE

Federated Learning–Driven Trustful and Privacy-Preserving Service Composition in Virtual Hospitals

Hedi Hamdi[1,*], **Zaki Brahmi**[2], **Sabeur Lajili**[3] **and Nabil Almashfi**[4]

[1]Department of Computer Science, Jouf University, Sakaka, Saudi Arabia
[2]Department of Software & IT Engineering, ETS Montreal, University of Quebec, Montreal, QC, Canada
[3]Department of Computer Science, ISITCOM, University of Sousse, Sousse, Tunisia
[4]Department of Software Engineering, Jouf University, Sakaka, Saudi Arabia

*Corresponding Author: Hedi Hamdi. Email: hhamdi@ju.edu.sa

Received: 02 April 2026; Accepted: 25 June 2026; Published: 28 September 2026

ABSTRACT: The increasing adoption of virtual hospitals and Internet of Medical Things (IoMT)-driven clinical workflows demands service composition mechanisms that can operate reliably under strict privacy constraints and heterogeneous provider ecosystems. Traditional trust-based composition approaches, such as majority voting or Bayesian aggregation, struggle to remain robust in large-scale, non-Independent and Identically Distributed (IID), and privacy-sensitive environments. This paper addresses these limitations by introducing PSCP-FL, a Federated Learning–Driven Privacy- and Trust-Aware Service Composition Framework for cloud- and edge-enabled virtual hospitals. We first define a comprehensive set of medical service trust attributes and construct a Federated Learning–based Trust Network (FLTN) that predicts provider trust while preserving data locality. FLTN enables the composer to prune untrustworthy services and select optimal compositions using a federated, dynamically updated evaluation model. Experiments demonstrate that PSCP-FL outperforms the Majority Voting and Bayesian Average-Based Trustful Service Composition (MVBA-TSC) baseline in standard deviation reduction (up to 40%), trust accuracy, robustness under non-IID and adversarial conditions, and execution time (15%–30% faster), while producing more stable and homogeneous service chains. These results confirm that PSCP-FL provides a reliable and privacy-preserving foundation for the composition of critical virtual hospital services.

KEYWORDS: Federated learning; service composition; srust management; srivacy preservation; sirtual hospitals; internet of medical things (IoMT)

1 Introduction

The digital transformation of healthcare has profoundly reshaped the way medical services are delivered. In recent years, the emergence of virtual hospitals has marked a turning point in this evolution, enabling the provision of specialized medical expertise and services remotely through interconnected cloud and edge infrastructures. Unlike traditional hospitals, which rely exclusively on physical co-location of staff and patients, virtual hospitals leverage advanced technologies to extend care beyond physical walls, offering patients continuous monitoring, remote diagnosis, and emergency support. This transformation has been accelerated by the proliferation of the Internet of Medical Things (IoMT), which integrates wearable sensors, bedside monitors, and smart diagnostic devices into unified healthcare ecosystems. As a result, clinical environments now generate unprecedented volumes of real-time data that can be used to enhance decision-making, reduce intervention delays, and improve patient outcomes.

Copyright © 2026 The Authors. Published by Tech Science Press. This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The rapid growth in the number and diversity of IoMT devices has also triggered the development of sophisticated service-based applications capable of exploiting distributed computing paradigms such as cloud and edge computing. These applications are essentially the result of the interaction and composition of multiple heterogeneous services, ranging from diagnostic AI systems to communication platforms and emergency dispatch tools. Such service compositions promise scalability, flexibility, and adaptability, which are essential properties in healthcare settings where patient states can evolve quickly and unpredictably. However, these benefits come with significant challenges. Since service interaction requires the collection, integration, and communication of sensitive health data, concerns about confidentiality, integrity, and regulatory compliance become paramount. Users and providers alike worry about the safety of private information, and any compromise in data security could severely undermine trust in virtual hospital infrastructures [1]. Thus, the dual challenge of ensuring privacy while maintaining reliable service composition has emerged as a central issue in the design of healthcare service systems.

At the core of this challenge lies the need for a trustful automatic service composition framework capable of generating feasible and efficient compositions under real-time constraints. In virtual hospital scenarios, where patients' lives can depend on the timely coordination of heterogeneous services, it is not enough to guarantee functional interoperability. Instead, the composer must dynamically manage what we define as a Trust Service Provider Network (TSPN), ensuring that only reliable providers are selected to participate in service composition. Within this network, the composer must evaluate the providers' trust scores, derived from historical reliability and recent performance, and select the most reliable path while jointly considering both Quality of Service (QoS) and the trust dimensions. For example, in an Intensive Care Unit (ICU) setting, a composer may need to simultaneously coordinate a neurosurgical diagnostic AI, a nurse chatbot, and an ambulance dispatch service. To be effective, such composition must not only meet latency requirements but also guarantee that the services involved are trustworthy and compliant with privacy regulations. Failure to address these requirements can lead to delays, inconsistent decisions, or even violations of patient confidentiality, consequences that are unacceptable in life-critical environments.

In the literature, the trust-sensitive service composition problem remains relatively underexplored. We divide existing approaches into two primary categories: Extra-trust (service-to-user) and less-explored Intra-trust (service-to-service). Extra-trust methods focus on reliability and QoS using techniques like Fuzzy Comprehensive Evaluation and weighted trust merging to assess service reliability [2], and multi-dimensional trust to protect against malicious providers [3]. Other Extra-trust work consider probability prediction [4] and privacy preferences [5,6]. Intra-trust compositions, emphasizing cooperation, use multi-step approaches based on social credibility [7] or Cartesian coordinate heuristics to select optimal cooperative pairs [8].

The study closest to ours is [9], which introduces trust-aware mechanisms where only trusted service providers join the composition. Trustworthiness is assessed via Bayesian averaging and majority voting. While these methods are simple and robust in homogeneous settings, they struggle in dynamic, adversarial, and multi-context environments. Recent work favors adaptive, reputation-weighted, and learning-based models that capture temporal drift and heterogeneous feedback reliability.

Federated learning (FL) has emerged as a promising paradigm for addressing privacy concerns by enabling collaborative model training without centralizing raw data. A growing body of literature explores FL in healthcare contexts, demonstrating its potential for medical image analysis, predictive modeling, and risk assessment [1,10,11]. However, these studies largely overlook how FL can be integrated into service composition frameworks. In particular, FL is rarely combined with trust-based composition to support privacy-preserving and adaptive service coordination in virtual hospital environments.

To fill this gap, this paper introduces a novel Federated Learning-based composition framework that directly addresses the Privacy- and Trust-Aware Service Composition Problem (PSCP) in virtual hospitals. Our framework integrates trust management, domain-aware clustering, and federated aggregation into the composition pipeline, thereby ensuring that only trustworthy providers contribute to service composition while strictly preserving privacy. Unlike previous work that optimizes QoS in isolation or applies FL solely to predictive tasks, our approach unifies these dimensions to achieve efficiency, reliability, and compliance in the delivery of critical life care. By incorporating federated learning, the framework eliminates the need for raw patient data exchange, thus maintaining confidentiality. By embedding trust evaluation into service selection, it guarantees that unreliable or adversarial providers are excluded from orchestration. Finally, by supporting dynamic reconfiguration, it ensures that compositions remain adaptive to evolving patient states, a requirement uniquely critical in ICU scenarios.

The major contributions of this work are fourfold. First, we formally define the PSCP and provide a mathematical model that incorporates patient state, service attributes, trust scores, and privacy-preserving constraints. Second, we design a federated learning-enabled service composition framework that integrates trust-aware service selection and domain-specific clustering to optimize compositions across edge and cloud environments. Third, we demonstrate the applicability of this framework through a motivating ICU scenario, showing how adaptive composition improves safety and timeliness. Fourth, we advance the state of the art by bridging two previously separate research streams—trust-aware service composition and federated learning for healthcare privacy—into a unified framework tailored for virtual hospitals.

The remainder of this paper is organized as follows. Section 2 reviews the most relevant prior studies, positioning our contribution within the existing literature. Section 3 introduces the fundamental concepts of federated learning and its relevance to privacy-preserving collaboration. Section 4 formulates the Privacy and Trust-Aware Service Composition Problem (PSCP), illustrated through a motivating ICU scenario. Section 5 presents the proposed Federated Learning-based Trust-Aware Orchestration Framework, describing its architecture, trust network construction, and service composition process. Section 6 details the experimental methodology, datasets and evaluation scenarios and discusses the results obtained. Finally, Section 7 concludes the article and outlines the directions for future research.

2 Related Works

The challenge of trustworthy service composition has evolved in parallel with the growing complexity of cloud–edge ecosystems and it has been approached from multiple angles. These include extra-trust models capturing service–user relationships, intra-trust models describing provider cooperation, trust aggregation methods that stabilize uncertain evidence, and federated learning frameworks that integrate trust with distributed intelligence that preserves privacy.

Extra-trust Models

Early efforts to ensure trustworthy service composition relied mainly on extra-trust models, which evaluate reliability between service providers and users based on Quality of Service (QoS), feedback, or Service-Level Agreement (SLA) compliance.

Noor et al. [12] provided one of the first systematic surveys of cloud trust management, identifying policy-, reputation-, and evidence-based mechanisms. Their study revealed major fragmentation and a lack of adaptive integration across trust dimensions, leaving contextual dynamics largely unexplored.

Sidhu and Singh [13] proposed an improved Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) method for ranking cloud providers according to SLA and QoS compliance. While

offering objectivity, it assumes static service behavior and centralized evaluation, which limits scalability and real-time adaptability.

John and Singh [14] extended trust assessment through fuzzy inference, integrating qualitative and quantitative parameters such as privacy and security. However, their model depends on expert-defined rules and fixed weightings, reducing its responsiveness to evolving contexts. Similarly, Alhadithy and Al-Shargabi [15] used fuzzy rules for QoS-based web-service composition, yet their selection mechanism focuses on individual transactions and neglects longitudinal trust evolution and cross-domain feedback.

In the context of fog and IoT, Ghaleb and Azzedin [16] adopted subjective logic to manage uncertain trust through the fusion of recommendations. The approach improves resilience against dishonest nodes, but remains computation-heavy and weakly personalized. Mo et al. [17] introduced an Active and Verifiable Trust Evaluation (AVTE) scheme that actively probes IoT devices for behavioral verification; despite its accuracy, it incurs additional energy and network costs unsuitable for large-scale or latency-critical systems.

Overall, these frameworks advanced extra-trust modeling from static reputation metrics to reasoning-driven evaluations, but still struggle with adaptivity, scalability, and privacy-preserving learning—gaps that motivate our proposed approach, which integrates distributed, data-driven, and context-aware trust computation.

Intra-trust Models

Intra-trust models address trust relationships among services or agents, focusing on cooperation and interdependence within composite service environments. Louati et al. [7] developed an early multi-agent coalition model in which autonomous service providers form dynamic coalitions guided by mutual trust. Although socially aware, this framework relies on localized heuristics and direct feedback, which limits scalability and adaptation in large, evolving networks. Later research extended this direction to edge-cloud ecosystems, combining inter-service trust with quality-of-service (QoS) to ensure reliable and privacy-preserving compositions [8]. In particular, this earlier work laid the groundwork for modeling cooperation between services rather than solely between users and providers. However, its heuristic evaluation and static trust formulation remain insufficient to capture evolving, cross-domain, or transitive trust relationships in dynamic distributed systems.

More recently, Huo et al. [18] introduced TrustGNN, a graph neural network framework that models the propagative and composable nature of trust graphs. Despite its improved structural representation, TrustGNN depends on centralized training and extensive labeled data, constraining its use in privacy-sensitive or large-scale environments. In summary, intra-trust models have progressed from heuristic and agent-based reasoning to structural learning, yet they still lack adaptive, privacy-aware, and decentralized trust evolution, gaps that motivate the new federated and context-sensitive framework proposed in this study.

Trust Aggregation

Trust aggregation models aim to integrate heterogeneous and sometimes uncertain evidence such as Quality of Service (QoS) indicators, user feedback, and interaction histories, into a unified reliability score. Bayesian methods have long provided a theoretical basis for such a fusion. Liu [19] proposed a Generic Bayesian Trust (GBT) framework unifying Beta, Dirichlet, and state-space models within a probabilistic perspective. These approaches quantify uncertainty and enable sequential updates of trust values, but they often assume evidence independence and rely on centralized computation, which restricts scalability and applicability in distributed environments. Hybrid aggregation strategies have subsequently emerged to improve resilience to biased or incomplete feedback. Sebri et al. [9] introduced the MVBA-TSC heuristic, which combines majority voting with Bayesian averaging to reinforce trust estimation in cloud-edge environments. This earlier work established the foundation for probabilistic trust fusion in sensitive service

composition. However, the model remains constrained by static weighting, limited contextual adaptation, and a dependence on network topology.

Overall, current aggregation approaches still face challenges related to contextualization, decentralization, and privacy preservation, motivating the development of a federated, adaptive, and privacy-aware aggregation framework for trustworthy service composition.

Federated Learning for Trustful Composition

As virtual hospitals and cloud-edge ecosystems expand, the limitations of classical trust models become increasingly evident. Centralized trust computation is no longer feasible since healthcare institutions must retain sensitive patient data locally in compliance with GDPR and HIPAA regulations. Trust evaluation must also be adaptive and evolve with the rapidly changing conditions of intensive care units and distributed healthcare environments. Furthermore, traditional approaches rarely address adversarial risks such as data poisoning, malicious providers, or heterogeneous (non-IID) data distributions across participating institutions. Federated Learning (FL) has emerged as a transformative paradigm to address these constraints. FL enables multiple institutions or edge nodes to collaboratively train global models without sharing raw data, thereby preserving local privacy. McMahan et al. [20] established the foundations of FL with the *FedAvg* algorithm, and the healthcare studies by Rieke et al. [21] and Sheller et al. [22] demonstrated its feasibility for multi-institutional collaborations in diagnosis, imaging, and prediction of chronic diseases. More recently, Bai and Dong [23] introduced an FL-based trust prediction framework for cloud and edge environments, showing that distributed trust estimation can enhance reliability while maintaining data privacy. Despite these advances, challenges persist. In fact, most FL frameworks assume honest participation, while in practice, contributors may behave unreliably or even maliciously. In addition, non-IID data across hospitals or service nodes complicates global model convergence and fairness. These issues are particularly critical for trustful service composition, where adversarial behavior or low-quality contributions can have direct and severe consequences. Thus, federated learning provides the foundation for *privacy-preserving, adaptive, and distributed trust computation*, but it must be extended with robustness and trust-awareness mechanisms tailored to sensitive service ecosystems.

Recent studies have increasingly investigated adversarial robustness and trust-aware federated learning in distributed intelligent systems. Recent work on enhanced model poisoning attacks and multi-strategy defense mechanisms demonstrated that conventional federated aggregation remains vulnerable to carefully crafted malicious updates and coordinated adversarial behavior [24]. Similarly, TFL-DT introduced a trust evaluation framework for federated learning in digital twin mobile networks, emphasizing dynamic trust estimation and adaptive participant evaluation in distributed environments [25].

In healthcare-oriented federated learning, recent surveys highlighted the growing importance of privacy-preserving collaborative intelligence, adaptive trust management, and secure IoMT integration in distributed medical systems [26]. Furthermore, abnormal client detection and incentive-aware federated learning mechanisms were recently proposed to improve robustness against unreliable participants and unstable updates in collaborative environments [27].

Compared with these approaches, PSCP-FL specifically targets privacy-preserving healthcare service composition in virtual hospital ecosystems. Unlike traditional federated learning frameworks that focus primarily on aggregation robustness, PSCP-FL jointly integrates trust-aware orchestration, adaptive provider filtering, QoS-aware composition, and federated privacy preservation within a unified collaborative architecture. Furthermore, the proposed framework directly connects trust evolution with healthcare service reliability and provider selection, thereby supporting adaptive and trustworthy orchestration under heterogeneous IoMT conditions. To compare the proposed framework with recent trust-aware

and Byzantine-resilient federated learning approaches, Table 1 summarizes the main characteristics of representative methods.

Table 1: Comparison with recent trust-aware and robust federated learning frameworks.

Method	Trust-Aware	Byzantine Defense	Service Composition	Healthcare Focus	Adaptive Aggregation
FedAvg [20]	No	No	No	No	No
FLTrust [28]	Yes	Yes	No	No	Partial
TFL-DT [25]	Yes	Partial	No	No	Partial
Poisoning-Defense FL [24]	Partial	Yes	No	General	Yes
PSCP-FL	Yes	Yes	Yes	Yes	Yes

Compared with recent Byzantine-resilient and trust-aware federated learning approaches, PSCP-FL emphasizes adaptive trust-aware orchestration specifically tailored to heterogeneous healthcare service ecosystems. While existing Byzantine-robust FL methods mainly focus on aggregation robustness and malicious update filtering, the proposed framework additionally integrates dynamic provider trust evolution, QoS-aware service composition, and privacy-preserving orchestration. Furthermore, the combination of event-driven synchronization and domain-specific clustering contributes to reducing communication overhead while maintaining stable trust convergence under heterogeneous IoMT environments. In light of these research directions, it becomes clear that existing approaches, although valuable, remain fragmented. *Extra-trust* models have advanced the understanding of user-to-service reliability but rely on centralized feedback and often lack explicit privacy safeguards. *Intra-trust* models capture cooperation among providers yet overlook large-scale heterogeneity and the need for adaptive, privacy-preserving mechanisms. *Trust aggregation* methods such as Bayesian or voting-based fusion stabilize noisy evidence but remain essentially static and dependent on centralized data. Finally, *federated learning* has introduced promising avenues for distributed and privacy-aware computation, but most frameworks assume honest participation and neglect the specific constraints of service composition in sensitive domains such as healthcare.

Despite recent progress in trust-aware and adversarially robust federated learning, existing approaches generally focus on aggregation robustness or participant trust evaluation independently. Few studies jointly address adaptive trust-aware healthcare service composition, privacy-preserving orchestration, and dynamic provider reliability management within heterogeneous IoMT-enabled virtual hospital ecosystems. Building upon these insights, the present work proposes an integrated framework that unifies these perspectives. By combining the stability of aggregation techniques, the inter-provider reasoning of intra-trust models, and the privacy guarantees of federated learning, our approach aims to provide an adaptive, decentralized, and trustworthy orchestration process. Rather than replacing earlier models, it extends them by translating their core principles into a federated, context-aware architecture suited to modern cloud-edge healthcare ecosystems.

3 Background on Federated Learning

3.1 Overview

Federated learning (FL) [29], a distributed training method, differs from traditional centralized training by allowing users in various locations to collaboratively train machine learning models without sending personal data to a central server. This approach aims to protect sensitive data (e.g., sensitive data collected by healthcare and IoMT services, such as biometric data) by keeping data on the device itself, enabling users to benefit from well-trained models without compromising privacy. The rise of AI chipsets has enhanced

the computational capabilities of client devices, making on-device model training feasible and increasingly shifting the focus from central servers to terminal devices. By leveraging terminal device resources, FL reduces the risk of data leakage during transmission, as sensitive information remains on individual devices. With the proliferation of mobile and IoT devices, federated learning capitalizes on vast, decentralized datasets across numerous devices, maximizing valuable resources while prioritizing data privacy and security [30]. This shift in training paradigms is paving the way for new research directions and applications in artificial intelligence that emphasize privacy and decentralized collaboration. Therefore, we adopt FL to support privacy-preserving collaboration in sensitive service-computing environments.

In practice, FL generally refers to distributed machine learning across multiple clients. Given N clients (e.g., service provider) $SP = \{sp_i|1 \ldots N\}$, each sp_i retains its own local dataset D_i and local model M_i, the FL process follows three phases (see Fig. 1):

1. *Initialization*: the central server shares an initial model with all clients.
2. *Local training*: each $sp_i \in SP$ trains its local model using its local dataset D_i. Concretely, clients collaborate to learn the model parameters w by addressing the following optimization problem [31]: $argmin\mathcal{L}_{glob}(w) = \sum_{i=1}^{N} p_i \mathcal{L}\geqq(w, D_i)$. Where $\mathcal{L}_i(w, D_i)$ is the local loss of client i.
3. *Model aggregation*: The central server aggregates the collected local models, $M_1, \ldots, M_n$, into an updated global model, $M^{t+1} \leftarrow Agg(M_i^{t+1}; i \in \{1 \ldots N\})$, which is then used to replace each user's local model.

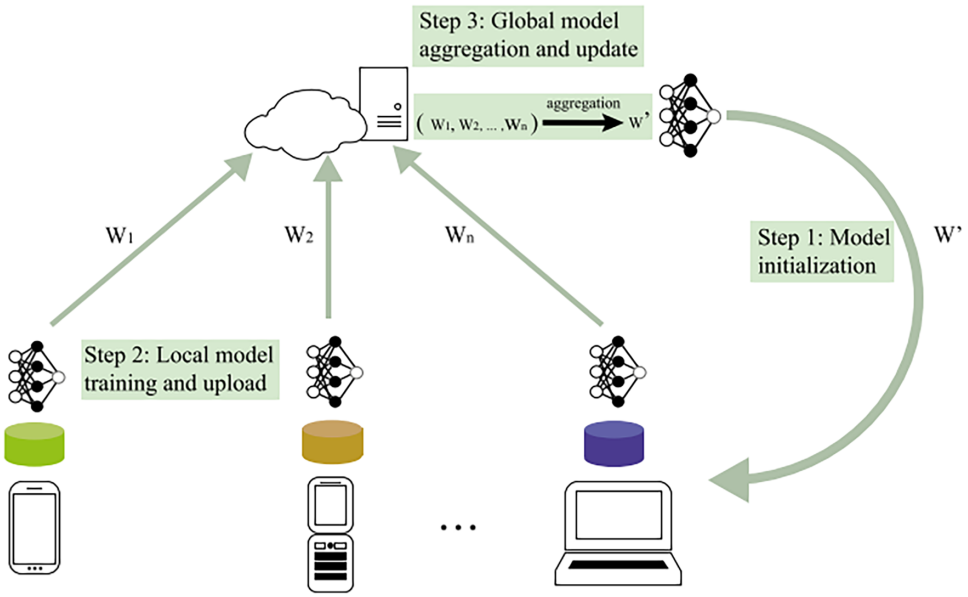

Figure 1: Federated learning enables clients to train locally and exchange only model updates with a central server, which aggregates them into a global model.

3.2 Model Aggregation

The main phase is model aggregation (MA), which can be described as the process of consolidating model parameters from all clients during each communication round to create an updated global model [32]. There are two types of MA: parameter-based aggregation and output-based aggregation. While parameter-based aggregation combines the trainable parameters of local models, output-based aggregation focuses on combining their output representations of the model (e.g., the output logits or compressed sketches) [33].

In the literature, many aggregation algorithms are developed such as FedAvg [20], FedProx [34], FedNova [34], Scaffold [35], MOON [36], Per-FedAvg [37], and FedAF [31] (it is not really a Model aggregation algorithm).

4 Problem Statement

Delivering care in virtual hospitals, and particularly in intensive care units (ICUs), requires composition mechanisms that are both dynamic and reliable. Patient states in these environments evolve rapidly, demanding immediate responses that may involve multiple heterogeneous services distributed across cloud and edge infrastructures. For example, a sudden spike in intracranial pressure (ICP) can require simultaneous activation of diagnostic AI services, conversational nurse agents, and ambulance dispatch systems. In such cases, any delay, inconsistency, or mis-coordination directly endangers patient safety. Traditional information systems in ICUs are not designed to support such complex and adaptive service composition, motivating the need for a more formal treatment of what we define as the Privacy- and Trust-Aware Service Composition Problem (PSCP).

4.1 Motivating Scenario: Adaptive ICU Orchestration

Consider a patient admitted to a virtual hospital ICU and continuously monitored by IoMT sensors measuring ICP, heart rate (HR), and oxygen saturation (SpO_2). At time t_0, the patient's ICP rises above 25 mmHg, a threshold associated with imminent cerebral damage. This event triggers the need for rapid and coordinated activation of services: (i) s_1, a neurosurgical diagnostic AI (sp_1) deployed at the edge, which analyzes ICP dynamics and CT images; (ii) s_2, a nurse chatbot (sp_2) in the hospital cloud, which interacts with caregivers to capture contextual information; and (iii) s_3, an ambulance dispatch service (sp_3) managed by the regional emergency network.

In conventional systems, such alerts are routed through rigid escalation protocols, where alarms appear on dashboards, nurses manually verify patient status, and emergency services are contacted via phone. This workflow is not only slow and fragmented but also fails to systematically evaluate the trustworthiness of service providers or safeguard patient data confidentiality. By contrast, PSCP envisions an orchestrator capable of dynamically composing these services in real time, while applying trust filtering and privacy-preserving mechanisms to ensure safety and compliance.

Fig. 2 illustrates this adaptive ICU orchestration. Patient data $p(t)$ is transformed into a query Q, which the orchestrator processes by evaluating candidate services s_1, s_2, s_3, each linked to a provider sp_i with trust score τ_i. Only services above the minimum trust threshold $\tau_{\min}$ are retained. Instead of exchanging raw data, providers train local models and send updates Δw_i, which are aggregated using federated learning into w_{t+1}. The orchestrator then computes a global recommendation $R(t)$—a risk classification, caregiver notes, and an emergency transfer instruction—delivered to the attending physician in near real time.

4.2 Problem Formalization

We formalize the Privacy-aware Service Composition Problem (PSCP) in edge–cloud settings as:

$$\text{PSCP} = (UQ, S, SP).$$

- $UQ = \{Q_i \mid 0 \le i \le q\}$: the set of user queries. Each query $Q_i = (Q_i^{\text{in}}, Q_i^{\text{out}}, Q_i^{\text{req}})$ is an abstract service specifying required *inputs*, expected *outputs*, and *requirements* (e.g., QoS/privacy). Inputs Q_i^{in} are the data provided by the user; outputs Q_i^{out} are the desired results; Q_i^{req} captures constraints such as latency, availability, cost, or privacy.

- $S = \{s_k \mid 0 \le k \le n\}$: the set of available services deployed by providers in SP. Each service $s_k = \langle s_k^{\text{in}}, s_k^{\text{out}}, s_k^{\text{nf}}, s_k^{\text{sp}} \rangle$ is described by its *inputs* s_k^{in}, *outputs* s_k^{out}, *non-functional* attributes s_k^{nf} (e.g., QoS/privacy guarantees), and the *provider* $s_k^{\text{sp}} \in SP$.
- $SP = \{SP_j \mid 0 \le j \le p\}$: the set of service providers. Each Service Provider SP_j hosts a subset $S_j \subseteq S$ and maintains a trust network $\text{SPTN}_j = \{\langle SP_\ell, TV_{j\ell} \rangle \mid SP_\ell \in SP\}$, where $TV_{j\ell}$ denotes the trust value assigned by SP_j to SP_ℓ. This value is predicted using Federated Learning, based on a set of features that includes: Response Time, Availability, Reliability, Usability, Credibility, Service Certification, Cost Satisfaction, and Prestige.

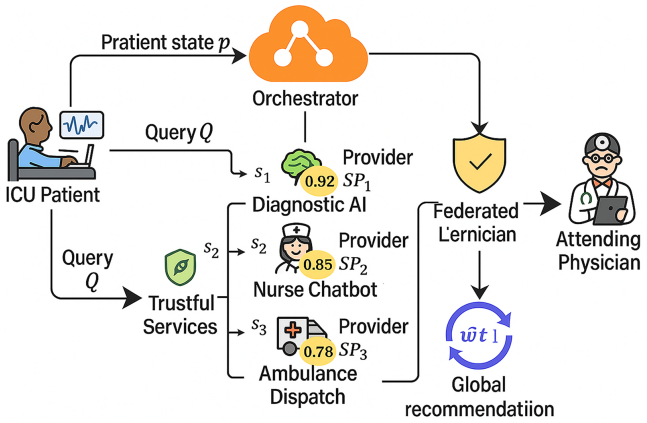

Figure 2: ICU workflow: patient state triggers a query; trusted services provide updates aggregated via FL to generate the global recommendation.

4.3 Service Trust Attributes for Virtual Hospital Composition

In the orchestration of virtual hospital services, trust evaluation ensures that only providers capable of meeting healthcare-critical requirements are selected. Trust attributes extend beyond simple QoS measures by integrating performance, usability, credibility, and security dimensions. Each parameter can be defined, evaluated, and formalized mathematically, enabling orchestrators to dynamically filter and compose services in a privacy-preserving and reliable manner.

Response Time

Response time measures how quickly a service reacts to a request. In a virtual ICU, a tele-neurosurgical AI that must interpret a rise in intracranial pressure (ICP) or an ambulance dispatch system that must confirm patient transfer requires ultra-low latency. It is evaluated by logging timestamps of request and response and computing the mean latency:

$$RT(s) = \frac{1}{N} \sum_{i=1}^{N} (t_i^{resp} - t_i^{req}).$$

For example, if an ICP monitoring device alerts at t^{req} and the diagnostic AI replies at t^{resp}, the difference quantifies responsiveness. Services with consistently low $RT(s)$ values are favored, particularly in emergency workflows. Recent studies confirm the importance of minimizing response time in federated healthcare orchestration [22,38].

Availability

Availability quantifies the extent to which a service remains accessible. In tele-ICU monitoring or virtual ward services, downtime can have catastrophic implications. Availability is measured by the ratio of uptime to total time:

$$Avail(s) = \frac{Uptime(s)}{Uptime(s) + Downtime(s)} \times 100\%.$$

For example, a patient monitoring service accessible 99.95% of the time is more trustworthy than one with frequent outages. Logs, probes, and SLA monitoring provide data for evaluation. High availability remains a cornerstone for dependable virtual hospitals.

Reliability

Reliability measures whether a service consistently provides correct outputs. In practice, a FHIR interface must always return valid patient bundles or a diagnostic API must produce valid classifications. Reliability is defined as:

$$Rel(s) = \frac{Successful\_Requests(s)}{Total\_Requests(s)}.$$

If a radiology AI receives 1000 image requests and 990 succeed, then $Rel(s) = 0.99$. Evaluators combine automated validation with fault-injection experiments to measure robustness. Reliability remains a critical QoS dimension in medical orchestration [39].

Usability

Usability reflects how effectively clinicians and patients can interact with a service. For example, a nurse chatbot designed for pain assessment must be intuitive and fast. Usability can be quantified by combining task success and user satisfaction:

$$Usab(s) = \frac{Successful\_Interactions(s)}{Total\_Interactions(s)} \times \frac{Satisfaction\_Score(s)}{MaxScore}.$$

Evaluation is performed using synthetically generated healthcare datasets designed to emulate realistic virtual hospital environments. Services with high $Usab(s)$ encourage adoption and improve trust in orchestration.

Credibility

Credibility refers to whether a provider honors commitments, including SLAs and regulatory promises. For example, an AI service claiming GDPR compliance must demonstrate adherence in audits. It is formalized as:

$$Cred(s) = \frac{Fulfilled\_Commitments(s)}{Declared\_Commitments(s)} \times ComplianceFactor(s).$$

Evaluation involves auditing SLA adherence, regulatory certifications, and post-incident analyses. In healthcare, credibility assures clinicians that providers are transparent and accountable.

Service Certification

Certification provides external proof of trustworthiness. Examples include FDA approval for diagnostic AI systems and ISO/IEC 27001 certification for data security. The certification score is computed as:

$$Cert(s) = \frac{\sum_{j=1}^{m} w_j \cdot c_j(s)}{\sum_{j=1}^{m} w_j},$$

where $c_j(s)$ indicates whether certification j is present. Services with stronger certification portfolios are preferred for inclusion in critical healthcare workflows.

Cost Satisfaction

Cost satisfaction balances value with expenditure. For example, a monitoring service that reduces nurse workload significantly justifies higher cost. It is computed as:

$$CS(s) = \frac{PerceivedValue(s)}{Cost(s)} \times NormalizationFactor.$$

Evaluation involves clinical pilots measuring time savings or improved outcomes against cost. Services with $CS(s) > 1$ are considered cost-effective and suitable for long-term adoption [39].

Deserve or Prestige

Prestige reflects the reputation of a provider or service. For instance, a diagnostic AI validated by a top-ranked hospital or cited in peer-reviewed trials carries more prestige. It can be expressed as:

$$Prestige(s) = \frac{\sum_{k=1}^{p} w_k \cdot rep_k(s)}{\sum_{k=1}^{p} w_k}.$$

Evaluation includes peer-reviewed validations, hospital rankings, and consortium memberships. Prestige is an indirect but powerful factor influencing service trustworthiness [22,38].

Trust (Composite Index)

The orchestrator must integrate all attributes into a unified trust score for decision-making. A composite trust index is computed as:

$$Trust(s) = \sum_{i=1}^{n} \alpha_i \cdot M_i(s), \quad \sum \alpha_i = 1,$$

where $M_i(s)$ are normalized attribute values. Weight assignments reflect clinical priorities (e.g., reliability for ICU alerts, usability for patient-facing chatbots). Evaluation combines metrics into a transparent scoring model [39]. To improve adaptability under dynamic healthcare environments, the weighting coefficients α_i are not maintained as static values. Instead, PSCP-FL dynamically adjusts trust-attribute importance according to observed provider behavior, workflow criticality, and federated performance evolution. Attributes exhibiting stronger correlation with prediction stability and service reliability receive progressively larger aggregation weights during successive federated rounds. This adaptive weighting strategy enables the trust model to better handle concept drift, heterogeneous provider behavior, and evolving QoS priorities in virtual hospital ecosystems. To further improve adaptability, the weighting coefficients are continuously recalibrated according to the observed contribution of each trust attribute to prediction stability and aggregation consistency. Let $Perf_i^{(t)}$ denote the contribution of attribute i during federated round t. The adaptive weight update mechanism is defined as:

$$\alpha_i^{(t+1)} = \frac{\alpha_i^{(t)} \times Perf_i^{(t)}}{\sum_{j=1}^{n} \alpha_j^{(t)} \times Perf_j^{(t)}}$$

where $Perf_i^{(t)}$ reflects the influence of attribute i on global trust prediction accuracy, aggregation stability, and service reliability during the current federated round. Consequently, attributes contributing more effectively to stable and accurate orchestration receive progressively larger weights, whereas attributes associated with unstable or less informative behavior gradually lose influence.

This adaptive weighting mechanism enables PSCP-FL to dynamically adjust trust evaluation according to evolving provider behavior, heterogeneous QoS conditions, and changing clinical priorities. In particular, healthcare-critical attributes such as reliability, security, and response time automatically gain higher importance during emergency or unstable operational conditions.

Security

Security ensures that sensitive patient data remains protected from attacks. For example, a federated learning service securing ECG data across hospitals must withstand intrusion attempts. Security is expressed s:

$$Sec(s) = \frac{DetectedThreats(s) - SuccessfulBreaches(s)}{DetectedThreats(s)}.$$

Evaluation involves penetration testing, SIEM monitoring, and regulatory alignment (HIPAA, GDPR). Services with $Sec(s) \approx 1$ demonstrate resilience, a non-negotiable attribute in healthcare [40].

5 Federated Learning-Based Trust Network of Service Providers (FLTN)

We start from the premise: *a service composition is trustworthy if its constituent services are operated by providers that mutually trust one another*. Building on this premise, our research addresses two key challenges: first, how to construct and maintain a trustworthy network of service providers using machine learning; and second, how to ensure that the underlying ML model is continuously updated to handle concept drift.

To address these challenges, we propose a federated learning (FL)-based trust composition framework, which relies on a *trust network* established during the design phase and exploited dynamically at runtime. In the design phase, service providers participate in a FL process to build a distributed trust graph that quantifies the reliability and credibility among providers. This trust network serves as the foundation for the composition phase. At runtime, when a service composition request is issued, the framework performs *real-time composition* by selecting and orchestrating only those services whose providers exhibit strong mutual trust relationships in the pre-established network. This separation between the offline trust-building process and the online composition ensures both **privacy-preserving trust evaluation** and **efficient, trustworthy service selection**.

We tackle the second problem with an event-driven FL model as described in the Algorithm 1. Here, a transaction between two service providers SP_u and SP_v, acts as a trigger ($e = (SP_u, SP_v, \text{meta})$), prompting the client to start an FL round (line 2) for global model updates.

Algorithm 1: Event-driven continuous federated learning-service provider SP_u

Input: Initial global model θ^0; initial trust graph $\mathcal{G} = (V, E, w)$; selection threshold τ; round timeout $T_{\max}$
Output: Updated global models $\{\theta^r\}_{r\geq 1}$; updated trust graph $\mathcal{G}$

1 while *system is running* **do**
    `//(1) Listen for service-level transactions`
2 **if** *transaction event* $e = (SP_u, SP_v, meta)$ *completes* **then**
        `// (2) Start a new FL round r, scoped by the event`

(Continued)

Algorithm 1 (continued)

$r \leftarrow r+1$; $\mathcal{C}_r \leftarrow$ `SelectParticipants` $(\mathcal{G}, \tau, e)$;
// Trust-aware participant selection based on Section 5.1.2.
broadcast `TriggerRound` $(r, \theta^{r-1}, \mathcal{C}_r)$;
`// (3) Collect local updates with deadline`
initialize $\mathcal{D}_r \leftarrow \varnothing$;
while *time* $< T_{\max}$ *and pending clients exist* **do**
  if *provider* $i \in \mathcal{C}_r$ *submits* $(\Delta\theta_i, metrics_i, trust\_signals_i)$ **then**
    $\mathcal{D}_r \leftarrow \mathcal{D}_r \cup \{(\Delta\theta_i, \text{metrics}_i)\}$;
`// (4) Aggregate and publish`
SP_u receives models $\mathcal{D}_r$ and aggregates: $\theta^r \leftarrow \text{Aggregate}(\theta^{r-1}, \{\Delta\theta_i\}_{(\cdot)\in\mathcal{D}_r})$;
broadcast θ^r to $\mathcal{C}_r$ (and optionally to all providers);
`// (5) Each SP updates its trust graph`
foreach SP_i **do**
    Receive global model;
    Update its graph $\mathcal{G}_i \leftarrow$ `UpdateTrust` $(\mathcal{G}_i, \mathcal{D}_r)$;

Unlike conventional asynchronous federated learning approaches that continuously aggregate stale client updates without contextual synchronization, the proposed event-driven federated learning mechanism selectively triggers aggregation rounds according to clinically relevant service interactions and significant trust-state variations. This strategy reduces unnecessary synchronization overhead while preserving consistency between trust evolution and healthcare workflow dynamics. Furthermore, associating aggregation rounds with recent provider interactions limits the influence of stale updates during collaborative model construction.

5.1 Building the FL-Based Trust Network (FLTN)

To build a trust network, each service provider collaboratively contributes to the formation of a global trust model by training a local trust model on its own interaction data. Rather than sharing sensitive raw information, providers exchange only their locally trained model parameters or updates, which are securely aggregated at the server. This process enables each provider to estimate the trustworthiness of its peers and gradually build an autonomous, privacy-preserving trust network among service providers.

Specifically, the server role is *context-dependent*. When a service provider SP_i initiates an interaction (e.g., to select collaborators for a composition), it temporarily assumes the role of an *aggregation server* and requests its neighborhood N_i of providers to share *local model parameters or updates* rather than raw interaction data. Each neighbor $SP_j \in N_i$ acts as a client and sends its locally trained trust model parameters θ_j^t. The provider SP_i aggregates these models to obtain a personalized trust predictor that reflects its local objectives and prior beliefs, while preserving data locality and privacy.

This dynamic server–client switching enables each provider to build a personalized, federated, and privacy-preserving trust network without exchanging sensitive raw data.

Although the trust estimation process in PSCP-FL is primarily data-driven, the proposed framework does not interpret trust prediction as a direct causal inference mechanism. To reduce sensitivity to spurious correlations across heterogeneous providers, trust evaluation incorporates temporal consistency verification, multi-round behavioral stability monitoring, and adaptive aggregation validation during successive federated learning rounds. Consequently, trust evolution depends not only on isolated client updates but also on persistent reliability consistency over time. This design reduces the influence of transient anomalies and

unstable local behaviors while improving robustness under heterogeneous IoMT-enabled healthcare environments. Nevertheless, maintaining robust and stable trust estimation in federated healthcare environments remains challenging due to data heterogeneity, communication constraints, and non-IID client behavior. Federated Learning (FL) achieves good learning performance but faces significant challenges in practical applications. A key issue is data heterogeneity, where each client's data is unique and non-independent, leading to non-IID data [41]. This causes traditional FL methods to suffer from performance degradation. Additionally, FL incurs high communication costs between clients and the server. Large models with millions of parameters require substantial data transfer over limited-rate channels, creating heavy traffic that can congest networks and extend training times [42], which is particularly problematic for real-time systems. To further evaluate scalability under large IoMT-enabled virtual hospital environments, additional simulations were conducted by progressively increasing the number of participating federated providers and connected medical devices. For this, we motivated by the use of Clustered FL (CFL), our framework is inspired by ACFL for active client selection while introducing a different domain-specific clustering strategy [41]. An enhancement of the original CFL approach, which accelerates the training process for non-IID data while simultaneously reducing communication costs. ACFL achieves these goals by efficiently clustering clients in each training round, leading to clients being directed toward more specialized models. Communication overhead is reduced by selecting the most informative clients (e.g., vote entropy, least confident) from each cluster using active metrics of each client. Hence, this method is most beneficial for enhancing the CFL model compared with a random selection strategy.

Prompted by the efficiency of ACFL, FLTN is a new method that enables each service provider to update its network by rejecting undesirable behaviors (outliers). In FLTN, each service provider, sp_i, can initiate the update process when an update to its trust network is required. sp_i trains its local model M_i using its local dataset, and in parallel sends local model requests to a set of selected service providers, once local models are received, sp_i runs the model aggregation and the outlier detection module to update its own trust network. Subsequently, the pipeline is broken down into the following tasks (Fig. 3):

1. Domain-specific clustering
2. Client selection
3. Trust network update

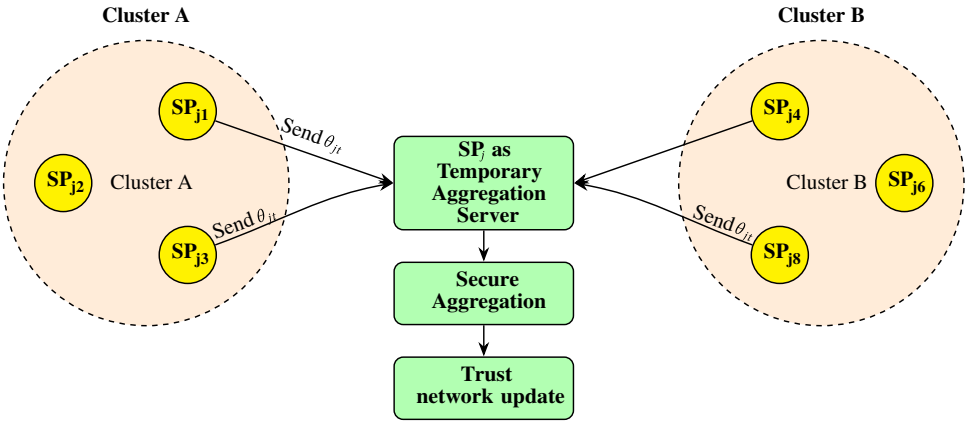

Figure 3: FLTN architecture.

5.1.1 Domain-Specific Clustering

In this paper, in contrast to ACFL, clustering clients/service providers is based on the active domain of clients (e.g., healthcare, finance, autonomous vehicles) rather than the similarities of client datasets distributions. The aforementioned, domain-specific clustering allows the FL model to capture the *high-impact QoS priorities* relevant to each sector, something that mere dataset similarities do not account for. For instance, in the healthcare domain, QoS metrics might prioritize response times and accuracy in critical diagnostics, while in finance, the focus could be on transaction processing speed and precision (e.g., fraud detection precision). By grouping clients based on these domain-specific QoS demands, the model can be tuned to prioritize metrics that align with the performance standards of each sector. This tailored approach not only enhances model relevance and performance for different service types but also supports compliance with regulatory standards. Thus, domain clustering provides a more contextually oriented approach that aligns the FL process with specific performance goals of each service domain, ultimately resulting in a more effective and relevant model.

During model aggregation, domain-specific clustering ensures that the aggregated models prioritize features and patterns relevant to each sector, reduce negative transfer (NT), and support efficient parameter updates. Clustering by domain minimizes performance degradation from "negative transfer" by keeping models within relevant QoS boundaries, preventing unrelated sectors from diluting each other's impact. Negative transfer refers to the undesirable reduction in learning performance in the target domain due to the influence of data or knowledge from the source domain. This issue has been a persistent and challenging problem in transfer learning (TL) [43]. Regarding parameter updates, it enables the aggregation function to focus on meaningful parameter adjustments for each sector's specific QoS needs. Furthermore, ref. [44] highlights the importance of domain-specific model updates to avoid biases during the global model aggregation due to variances in parameter importance and inconsistency in update directions.

5.1.2 Client Selection

Client selection (or sampling) plays a pivotal role in ensuring the efficiency, effectiveness, and fairness of the FL training process. The goal is to select the most representative and informative set of clients to participate in the FL training phase. However, it introduces several challenges. A major issue is the *non-IID (non-independent and identically distributed)* nature of client data [45], which can lead to biases in the global model if not managed properly. For instance, selecting clients with skewed or highly divergent data distributions can impede the convergence of the global model and degrade its overall performance [46]. Therefore, it's crucial to choose clients with representative, diverse, and high-quality data samples that improve the learning task [47]. Another statistical challenge is *concept drift*, which denotes a shift in the data distribution across different periods [48]. Indeed, client selection needs to consider the dynamic environment. *Heterogeneity* of client resources—such as varying computational power, memory, and network capabilities—complicates the selection process, as does the risk of client dropout due to unreliable participation. *Fairness* is another critical concern, as repeated selection of certain clients can introduce bias and reduce the representativeness of the model [49]. *Communication* overhead is also significant, requiring strategies to minimize data transfer while maintaining performance [50]. Finally, *privacy and security* concerns must be addressed to protect client participation patterns and mitigate the risk of adversarial attacks. Together, these issues necessitate the development of robust, efficient, and fair client selection mechanisms to ensure the success of federated learning systems.

In the literature, two common strategies are [51] (a) local samples number-based selection [20], and (b) random selection [52], which may lead to slower convergence and poor model performance across devices [53]. Recent studies have proposed advanced sampling techniques that account for dynamic

system constraints, enhance the convergence speed of federated optimization, and improve model accuracy [41,51,53,54]. Similarly, aiming to tackle these challenges to improve the performance and reliability of FL systems, we opt for considering the client selection as an optimization problem to effectively balance the trade-offs between different objectives (e.g., fairness, privacy, resources). Motivated by the finding of [41], where client selection is based on a set of combined metrics (e.g., uncertainty sampling, query-by-committee (QBC), loss) for measuring the informativeness of each client, we propose a selection method combining a set of metrics. Therein, our selection is driven by the balance between ensuring Client Diversity (CD), Client Availability and Frequency of Updates (CAFU), Client Contribution (CC), Trustworthy Client (TC), and Model Contribution (MC).

Client Diversity (CD)

Client Diversity (CD), which refers to the diversity of local data [55] and/or gradients (or models) [51] across different clients in federated learning (FL), is a key characteristic of non-IID data. It helps prevent overfitting to specific data patterns and improves the model's ability to handle a variety of real-world situations. In general, the diversity strategy measures how well a selected subset of clients represents the overall client population when their updates are aggregated at the server. Furthermore, diversity-based client selection can help address fairness by ensuring that clients with diverse data distributions are included in the training process, thereby preventing bias and promoting a more equitable contribution to the global model. This objective is supported by the proposed domain-specific clustering of clients, which groups clients with similar data distributions while ensuring diverse and balanced participation in the FL process [46]. Client diversity can be quantitatively measured using the *Weight Divergence*, calculated as the normalized Euclidean distance between a client's locally trained model (w_i, b_i) and the global model (w_g, b_g):

$$\mathrm{Div}_i = \frac{\|w_i - w_g\|_2}{\|w_g\|_2}. \tag{1}$$

Large divergence values indicate that the client produces updates that deviate significantly from the common learning direction, potentially due to noisy, low-quality, or even adversarial behavior.

Trustworthy Clients' Selection (TC)

In the context of a service composition scenario, trust signifies the confidence a service provider exhibits in another's ability to perform designated tasks and act as a reliable collaborator. The objective of this strategy is twofold. Firstly, in our context and motivated by the finding of our previous work [9], a service provider p_i mainly tends to cooperate with trustworthy providers; hence, provider trust score plays a pivotal role in our solution. Secondly, in the context of FL, existing FL systems frequently exhibit vulnerabilities to attacks and face challenges in meeting security requirements in practical applications [56]. Each *worker* provider, p_j, has a set of associated trust score $ts_j = \{ts_j^i | 1 \ldots N-1\}$, which reflects its reliability and the risk of misbehaving. ts_j^i is trust score of the worker provider p_j computed based on the historical interaction tracked by p_i. It is crucial to highlight that, in addition to individual interaction with p_i, our approach incorporates peer interactions to determine ts_j^i. In fact, p_j must demonstrate consistent alignment with their peers. This expanded concept of trust fosters collaboration and synergy throughout the network. The main goal of selecting informative clients from each cluster is to identify those clients whose data best captures the key patterns and characteristics of the entire cluster and to ensure efficient and effective model aggregation [41].

Model Contribution (MC)

This metric allows the server to select clients based on their effective contribution to the global model. In FL, determining which client model contributes the most to enhancing the global model involves evaluating the impact of individual client updates [56]. Hence, the model contribution of a client measures how much its update improves the global model's performance. For linear regression with mean squared error (MSE) loss, the contribution of client k at round t can be computed as the change in validation loss when applying its update Δw_k to the current global model w_t:

$$C_k = \mathrm{MSE}(X_V w_t,\ y_V) - \mathrm{MSE}(X_V(w_t + \eta\,\Delta w_k)\,,\ y_V)\,,$$

$$\mathrm{MSE}(\hat{y}, y) = \frac{1}{|V|}\| \hat{y} - y \|_2^2, \quad (2)$$

where (X_V, y_V) denotes a small validation dataset and η is the learning rate. A positive C_k indicates that the client's update improves the global model, while a negative value suggests it degrades performance.

Aggregation

Selecting a subset of clients in each cluster to participate in federated learning (FL) training is a challenging task, as it must simultaneously consider Client Diversity (CD), Client Availability and Frequency of Updates (CAFU), Client Contribution (CC), Trustworthy Client (TC), and Model Contribution (MC). Hence, selecting a subset of clients within each group, constitutes a 0–1 integer programming problem with vector weight constraints. This problem is known to be NP-complete, as proven by Li et al. [55]. Furthermore, the aggregation weights are dynamically updated during successive federated rounds according to client contribution stability, trust evolution, and global prediction consistency.

To aggregate the client-selection metrics—CD, CAFU, TC, MC, and CC-We define a weighted aggregation function, $F_{\text{aggregate}}$ (Eq. (3)), that reflects the importance of each strategy in the context of the proposed framework.

$$F_{\text{aggregate}} = w_1 \cdot \mathrm{CD} + w_2 \cdot \mathrm{CAFU} + w_3 \cdot \mathrm{CC} + w_4 \cdot TC + w_5 \cdot MC \quad (3)$$

where α, β, and γ represents the weights assigned to each strategy. To resolve this minimization optimization problem, the Quantum-inspired Gravitational Search Algorithm (EQ-GSA), proposed by [57], is employed.

5.2 Threat Model and Adversarial Defense Mechanism

PSCP-FL considers three categories of adversarial behavior that may arise in distributed virtual hospital environments: model poisoning attacks, Byzantine updates, and malicious client collusion.

In model poisoning attacks, compromised providers intentionally manipulate local model updates to degrade the global federated model or bias trust prediction outcomes. Byzantine participants may generate arbitrary, inconsistent, or noisy updates due to compromise, malfunction, or malicious intent. In collusion attacks, multiple malicious providers cooperate to artificially increase their trustworthiness or negatively influence the trust evaluation of legitimate participants.

To improve robustness against such threats, PSCP-FL incorporates a trust-aware filtering and adaptive aggregation mechanism. Each local update is evaluated according to historical trust evolution, behavioral consistency, aggregation deviation, and service reliability indicators. Providers whose updates repeatedly

diverge from expected collaborative behavior undergo progressive trust decay, reducing their influence during subsequent aggregation rounds.

Furthermore, aggregation weights are dynamically adjusted according to continuously updated trust scores, thereby limiting the impact of unreliable or adversarial participants. Providers exhibiting persistently abnormal behavior are excluded from future federated aggregation rounds and service composition procedures.

Unlike conventional federated learning aggregation methods that equally weight all clients, the proposed mechanism introduces adaptive trust-guided aggregation specifically tailored to safety-critical healthcare environments.

Although the proposed defense strategy significantly improves robustness against poisoning and Byzantine behavior, highly adaptive adversaries capable of mimicking legitimate update distributions may still partially evade trust filtering mechanisms. Moreover, PSCP-FL combines temporal trust evolution with multi-round behavioral consistency verification, thereby reducing the probability that short-term adversarial mimicry can successfully maintain high trust scores over prolonged federated interactions. Investigating advanced anomaly detection and causality-aware trust modeling for adaptive adversarial environments remains part of our future research directions.

5.3 Trust-Aware Service Composition

The service-composition mechanism adopted in our framework is inspired by the *MVBA TSC* model [9], in which one of the present authors participated. In that earlier approach, the orchestrator constructed an end-to-end composition by interconnecting local processes proposed by trusted providers belonging to a Trust Service Provider Network (TSPN). Each provider advertised its functional capabilities, Quality-of-Service (QoS) parameters, and trust level. The orchestrator then selected the most appropriate services and combined them into a coherent execution plan that maximized both technical performance and social reliability.

In the present work, we extend this principle by integrating the trust evaluation produced by the Federated Learning Trust Network (FLTN). The goal is to allow trust information to evolve dynamically and privately as hospitals and edge devices update their local learning models. When a user query Q_u is submitted, the orchestrator performs a sequence of coordinated operations that ensure the resulting composition remains trustworthy, privacy-preserving, and adaptive to current conditions.

First, the orchestrator retrieves the most recent trust scores τ_i from the FLTN. Any provider whose trust value falls below the acceptance threshold $\tau_{\min}$ is excluded, together with its associated services. The Service Class Network (SCN) is therefore pruned to contain only candidates considered reliable according to the latest federated assessment.

Second, the orchestrator identifies the start class C_s whose input signature matches the input of the query. For each candidate service $s_i \in C_s$, The evaluation score combines four complementary criteria: the node degree, the average trust value, the QoS score, and the provider trust score. The global evaluation function,

$$Eval(s_i) = \frac{deg(SP(s_i)) + \text{avg}(TV(SP_s)) + QoS_i + \tau_{SP}}{3}, \tag{4}$$

yields a balanced indicator that captures both objective performance and collective reputation. The service with the highest score is selected; its output becomes the input of the next class C_{i+1}, and the process is repeated iteratively until a service generates an output matching $out(Q_u)$. The sequence of selected services

defines the final composition plan $Comp^*(Q_u)$, while the global trust index of the composition, $\bar{\tau}$, is obtained as the mean of the trust scores of the participating providers.

In our federated extension, the orchestrator does not rely on static trust tables but instead queries the FLTN for continuously updated values of τ_i predicted by local models. Each update round refines these values based on local experience and QoS feedback, without exposing sensitive data to external entities. Consequently, the composition process adapts in real time to variations in provider behavior, workload, and reliability while preserving the confidentiality of hospital data. Although trust scores guide provider selection and adaptive aggregation during orchestration, the evaluation of PSCP-FL is not restricted to trust-oriented metrics alone. To avoid circular validation, the experimental assessment additionally relies on independent system-level and prediction-oriented indicators, including RMSE, R^2 score, convergence stability, execution time, communication overhead, and prediction accuracy. These metrics are computed independently from the trust optimization process and therefore provide an external validation of the effectiveness and robustness of the proposed framework.

To illustrate the process, consider a request in a virtual Intensive Care Unit (ICU) where a patient exhibits a sudden increase in intracranial pressure. The composer formulates the query

$$Q_u = \big(\{ICP(t),\ HR(t),\ SpO_2(t)\},\ \{Diagnosis,\ Alert,\ TransferPlan\},\ Req = \{Latency < 3s,\ GDPR,\ Reliability > 0.9\}\big). \quad (5)$$

The SCN then identifies three relevant service classes: C_1 for neurosurgical diagnostic AI services, C_2 for nurse-assistant chatbots handling patient interaction, and C_3 for emergency dispatch systems. Using the most recent trust and QoS information, the orchestrator evaluates each candidate and selects the best combination: s_1 from C_1, s_4 from C_2, and s_6 from C_3, forming the optimal composition

$$Comp^*(Q_u) = \{s_1 \to s_4 \to s_6\}, \qquad \bar{\tau} = 0.93.$$

The resulting orchestration coordinates an AI-based diagnosis, a conversational monitoring service, and a secure ambulance dispatch—ensuring timely, trustworthy, and regulation-compliant clinical support.

Compared with the original MVBA TSC mechanism, this federated version provides an adaptive, data-driven evolution of trust across distributed providers. All trust computations occur locally, and only model parameters are aggregated through the FLTN, guaranteeing compliance with privacy regulations such as GDPR and HIPAA. The resulting service composition is therefore both technically efficient and socially reliable, offering a robust foundation for autonomous orchestration within virtual-hospital ecosystems. To better illustrate the operational logic of the proposed approach, Algorithm 2 presents the detailed procedure followed by the orchestrator to construct a federated, trust-aware, and privacy-preserving service composition.

Algorithm 2: Trust-aware federated service composition

Input: User query $Q_u = \langle in(Q_u), out(Q_u), Req(Q_u)\rangle$;
Service Class Network SCN; Provider set $\mathcal{SP}$;
Federated trust model f_θ; Threshold $\tau_{\min}$.

Output: Final composition $Comp^*(Q_u)$ and global trust $\bar{\tau}$.

(Continued)

Algorithm 2 (continued)

Step 1: Trust Refresh
foreach *provider* $SP_j \in \mathcal{SP}$ **do**
  Compute local trust score $\tau_j \leftarrow f_\theta(x_j)$;
  Send model update to FLTN for secure aggregation;
end
Step 2: Pruning of Untrusted Services
foreach *service* s *in* SCN **do**
  if $\tau_{SP(s)} < \tau_{\min}$ **then**
    Remove s from SCN;
  end
end
Step 3: Initialization
Find start class C_s such that $in(C_s) = in(Q_u)$;
Initialize composition plan $\pi \leftarrow \langle\rangle$;
Set current input $x \leftarrow in(Q_u)$;
Step 4: Iterative Selection
while $x \neq out(Q_u)$ **do**
  Identify current class C_k with input x;
  foreach *candidate service* $s_i \in C_k$ **do**
    Compute evaluation score: $Eval(s_i) = \dfrac{deg(SP(s_i)) + avg(TV(SP_s)) + QoS_i + \tau_{SP(s_i)}}{4}$;
  end
  Select best service: $s_k^* \leftarrow \arg\max_{s_i \in C_k} Eval(s_i)$;
  Append s_k^* to π;
  Update data flow: $x \leftarrow s_k^*.output$;
end
Step 5: Trust Aggregation
Compute global trust: $\bar{\tau} = \frac{1}{|\pi|} \sum_{s_i \in \pi} \tau_{SP(s_i)}$;
Return $(Comp^*(Q_u) = \pi, \bar{\tau})$;

6 Evaluation and Experimental Analysis

6.1 Experimental Setup

The experiments were conducted using a lightweight federated learning framework implemented in Python, utilizing standard scientific libraries such as *NumPy* and *scikit-learn*. A linear regression model was employed due to its interpretability and suitability for trust prediction. The federated learning configuration was selected to ensure stability and fairness across heterogeneous client datasets. Specifically, 10 federated rounds were used to allow adequate convergence of the global model; a batch size of 32 was chosen to balance gradient estimation noise and computational efficiency; a learning rate of 0.01 was adopted to provide stable and gradual weight updates; and 5 local epochs per client were employed to limit client-side model drift while ensuring meaningful local training. Model performance was assessed using RMSE and R^2, whereas client behavior was monitored through trust scores and neighbor dynamics. Complete experimental source code is publicly available on request GitHub.

6.2 Dataset Design and Generation

The specialized nature of the Trust Service Composition Problem and the absence of a large, publicly available dataset required us to leverage publicly available QoS datasets for service composition, mainly the QWS dataset (v1 and v2) and the dataset in [58]. We then augmented these existing data by using Generative AI (e.g., GPT-4) to define and enrich the crucial trust features, thereby creating a comprehensive dataset that integrates both quantitative quality metrics and subjective trust properties for reliable analysis. The developed prompt is mainly based on existing market reports and academic literature [8,9,59–61]. The final dataset was programmatically generated using Python, drawing numerical values from a normal distribution to simulate natural variability. We ensured the dataset's reliability through a three-pronged validation process:

- Statistical Checks: We compared summary statistics (mean, variance) and conducted distributional analyses against industry trust benchmarks to ensure the data were plausible.
- Manual validation: Authors and domain experts reviewed the feature values and their relationships, confirming the dataset's alignment with real-world characteristics of service provider trust and behavior.
- Sensitivity Analysis: We systematically varied key parameters (e.g., Certification, Credibility) and confirmed that the dataset produced robust, logical, and consistent outcomes across various service composition scenarios.

The generated dataset is composed of samples representing all the services provided by clients. Each client, representing a service provider, owns a local dataset describing various service quality attributes, including *ResponseTime (ms)*, *Availability (%)*, *ReliabilityScore*, *Usability (%)*, *Credibility*, *Certification*, *CostSatisfaction (%)*, *Prestige (1–5)*, and *Security (%)*. The target variable is the *TrustIndex*, a continuous score representing the overall trustworthiness of a service provider. To ensure realism and challenge the federated learning process, datasets were generated with heterogeneity across clients, simulating non-IID (non-identically distributed) conditions.

6.3 Realism and Limitations of the Simulation Environment

Although the current evaluation relies on synthetically generated healthcare service datasets, the simulation environment was designed to approximate realistic virtual hospital conditions by incorporating heterogeneous IoMT providers, non-IID data distributions, dynamic service availability, fluctuating QoS behavior, and variable trust degradation patterns. The generated scenarios emulate realistic healthcare orchestration constraints, including emergency response latency, provider instability, communication variability, and security-sensitive collaboration among distributed medical entities.

To improve realism, the simulation integrates heterogeneous provider behaviors across cloud-edge infrastructures and varying reliability profiles inspired by operational virtual hospital workflows. Moreover, adversarial and unstable participant behaviors were introduced to evaluate the robustness of the proposed framework under challenging collaborative conditions.

Nevertheless, synthetic datasets cannot fully reproduce the complexity of real-world clinical service logs, institutional workflow dependencies, and operational IoMT environments. Therefore, future work will extend PSCP-FL using real healthcare datasets and clinical IoMT traces obtained from operational virtual hospitals, subject to privacy regulations, ethical approvals, and institutional data-sharing constraints.

6.4 Feature Importance Interpretation in Provider-Centric Trust Modeling

The obtained feature importance values indicate that the most influential factors in determining trust among cloud providers are **Response Time**, **Reliability**, **Cost Satisfaction**, and **Security**. This finding aligns with the nature of inter-provider communication in composite services, where performance stability and secure service exchange are paramount. Response Time (0.57) dominates trust formation as it directly

affects end-to-end latency and user-perceived Quality of Service (QoS). Reliability (0.42) and Availability (0.27) further reinforce confidence by ensuring continuous service delivery and reduced failure risk. Cost Satisfaction (0.40) reflects the economic efficiency and fairness of collaboration, while Security (0.38) safeguards data integrity and confidentiality during cross-provider transactions. In contrast, Credibility (0.20) and Certification (0.16) have moderate impact, while Usability (0.12) and Prestige (0.07) contribute minimally. Fig. 4 illustrates the global feature importance scores obtained from the proposed federated learning framework, highlighting the relative contribution of each trust-related feature.

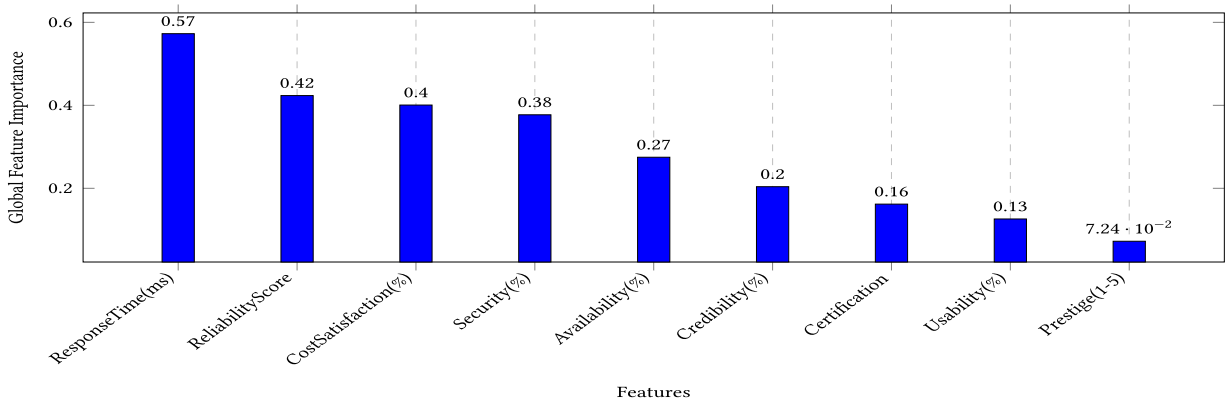

Figure 4: Feature importances in federated learning.

6.5 Federated Learning–Based Aggregated Model Construction

The experimental methodology is organized around scenario-based evaluations designed to examine key aspects of federated trust management among service providers. Each provider seeks to construct its own trust network to facilitate information exchange and collaboration in delivering composite services. This involves identifying and selecting the most efficient and trustworthy clients (i.e., other service providers) whose contributions enhance the aggregated federated learning model.

For illustration, we consider a system of 1000 service providers offering heterogeneous services, with historical interactions stored locally. Each scenario addresses a specific research question, emphasizing trust evolution, federated learning cooperation, and resilience to unreliable or low-quality clients. The results demonstrate the impact of these factors on both global model performance and the dynamics of client trust across federated rounds.

6.5.1 Scenario 1: Aggregated Trust-Based Model without Client Selection

In this scenario, clients possess heterogeneous dataset sizes ranging from 1000 to 37,500 samples. All clients start with an initial trust score above 0.5 and are considered trusted. To extend Scenario 1 to a realistic large-scale deployment, we simulate the trust-based federated training process with 1000 clients. Since visualizing all individual client behaviors is infeasible, we categorize clients according to their final trust contribution to the global model. Table 2 shows that:

(1) 60% of clients achieve high trust ($T > 0.70$), meaning they consistently improve the aggregated global model;
(2) 22% of clients fall into the moderate trust range ($0.50 \leq T \leq 0.70$), providing only partial or unstable benefit;
(3) 18% of clients exhibit low trust ($T < 0.50$), meaning their updates degrade the global model.

Table 2: Categorization of 1000 clients by final trust contribution.

Trust Category	Trust Range	Percentage of Clients	Avg. Trust
High contributors	$T > 0.70$	60%	0.78
Moderate contributors	$0.50 \leq T \leq 0.70$	22%	0.60
Low contributors	$T < 0.50$	18%	0.35

Fig. 5 visualizes this distribution, confirming that the trust mechanism scales effectively to large populations: the majority of clients (60%) are reliable contributors, while the remaining clients (40%) are either moderate or low contributors. These results highlight the necessity of isolating unreliable participants and retaining only beneficial contributors in subsequent training phases.

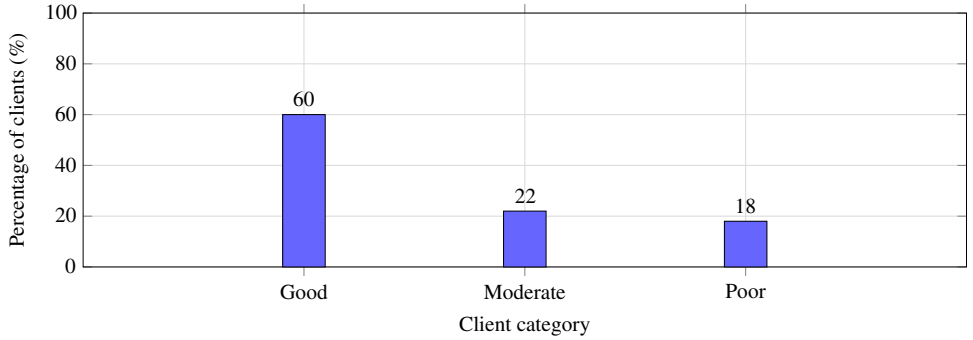

Figure 5: Distribution of 1000 clients by their impact on the global model.

Despite the heterogeneity in client contributions, the global model remains stable throughout the training process, with an RMSE of approximately 0.380 and an R^2 of approximately 0.79.

To visually illustrate the evolution of the training process and the client behaviors, we restrict this experiment to a representative subset of five clients, since plotting all 1000 clients is not feasible. Fig. 6 shows that trust scores increase monotonically for all clients, and a fully connected neighbor graph (client composition) emerges as early as Round 2. By Round 10, trust converges close to 1.0 for all clients (C2–C5), while C1 stabilizes around 0.90 due to its limited ability to contribute to the global model. The global model remains stable during the entire process (RMSE $\approx$ 0.380, $R^2 \approx 0.790$), confirming that the proposed trust mechanism ensures fair collaboration and robust convergence despite pronounced dataset-size heterogeneity and uneven local performance.

6.5.2 Scenario 2: Aggregated Trust-Based Model with Optimal Client Selection

In this scenario, we apply the client selection method introduced in Section 5.1.2, where client contribution, dataset diversity, and trust score serve as the primary criteria for categorization. Only clients whose updates (i.e., local model parameters) improve the global model quality are retained, while those whose contributions degrade it are excluded from the candidate set.

To identify representative clients, the efficiency of each client in improving the aggregated model is evaluated using the aggregation function (3). This function computes the *client reliability score*, which quantifies the positive or negative impact of each client on the global learning process. An increase in the client reliability score corresponds to a higher trust score for that client.

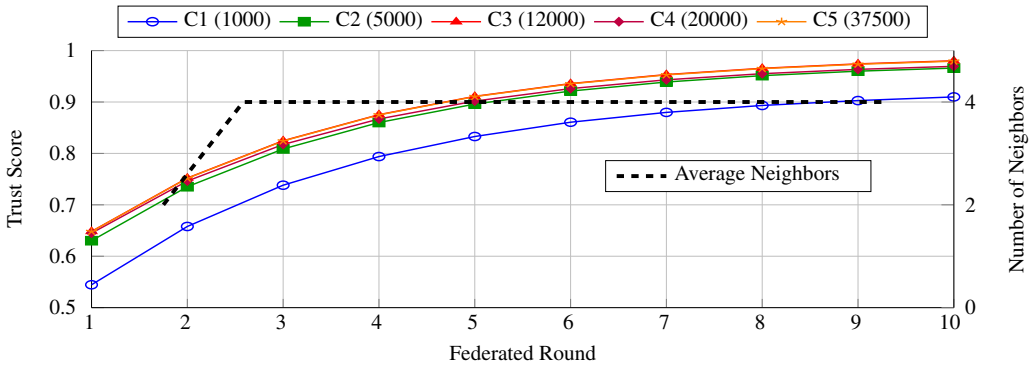

Figure 6: Scenario 1: Evolution of trust scores (solid lines) and number of neighbors (dashed line) across 10 federated rounds without client selection.

Experiment 1: Client Selection

The 1000 clients are categorized based on their trust scores (see Table 3), and only high-trust clients (50%) are retained. The aggregation function (3) computes each client's reliability score, reflecting its trust. High-reliability clients are selected as representatives, while moderate- and low-trust clients are excluded, ensuring that only the most beneficial participants contribute to the global model.

Table 3: Categorization of 1000 clients by final trust contribution using client selection.

Trust Category	Trust Range	Percentage of Clients	Avg. Trust
High contributors	$T > 0.75$	50%	0.85
Moderate contributors	$0.60 \leq T \leq 0.75$	25%	0.68
Low contributors	$T < 0.60$	25%	0.50

Table 4 shows that training with only the top 60% high-quality clients significantly improves the global model, reducing RMSE from 0.380 to 0.212 and increasing R^2 from 0.790 to 0.95. The average trust rises to 0.92, indicating that only the most reliable contributors participate.

Table 4: Global model performance using only high-quality contributors.

Metric	Scenario 1 (All 1000 Clients)	Scenario 2 (Top 60% Clients)
RMSE	0.380	0.212
R^2	0.790	0.95
Average Trust	0.85	0.92

Fig. 7 illustrates that trust converges faster and at higher values, confirming that selective aggregation of top contributors enhances both model quality and network reliability.

Experiment 2: Impact of Data Drift in Client Selection

This experiment investigates the effect of data drift (or outliers) on trust scores and the selection of reliable representative clients. To enable clear visualization, this experiment is limited to five clients. Outliers are introduced with moderate aggressiveness (outlier_fraction = 0.2, outlier_multiplier = 20). As shown

in Fig. 8, with a large outlier dataset, Client 4 significantly degraded the global model and was confirmed as a persistent outlier. The system tracked RMSE and R^2 histories over multiple rounds to dynamically adjust trust scores, detect under-performing clients, and eliminate persistent outliers from aggregation, ensuring robust global model updates. At Round 5, Client 4 is detected as a persistent outlier, leading to a sharp improvement in the global model's performance: the Global RMSE drops from approximately 0.36 to 0.10, and the Global R^2 increases from negative values (around −1.0) to positive (0.66) (see Fig. 9). This indicates that pruning the outlier significantly stabilizes and enhances federated learning aggregation.

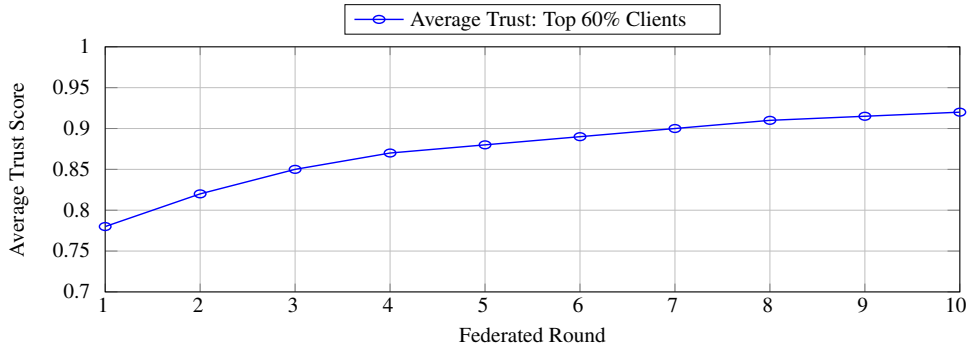

Figure 7: Scenario 2: Average trust evolution across 10 federated rounds for high-quality clients (top 60%) selected from Scenario 1. Trust converges faster and reaches higher values, leading to a more robust global model.

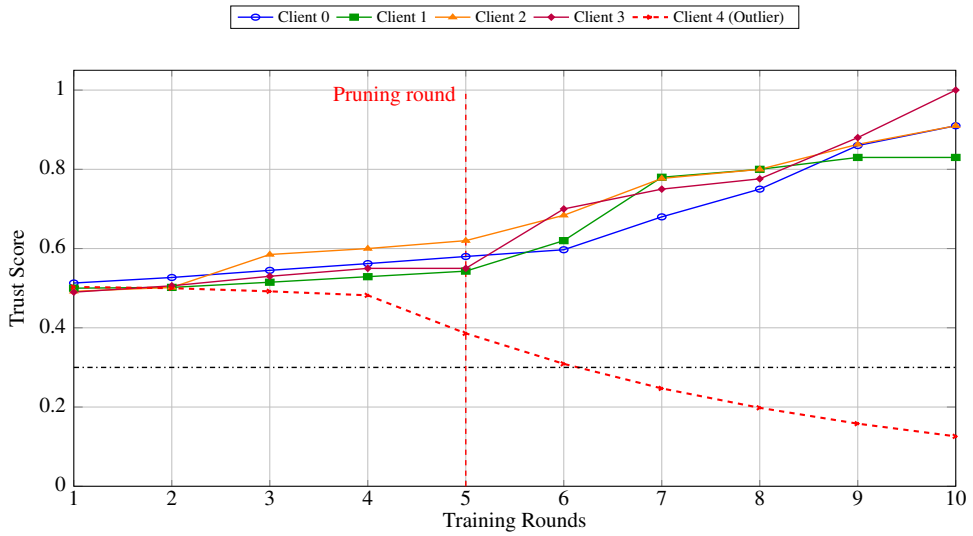

Figure 8: Exponential evolution of trust scores across federated rounds. Trusted clients exhibit rapid exponential growth, while client 4 with largest dataset, identified as a persistent outlier at round 5, is gradually pruned.

Experiment 3: Federated Training with Selected High-Trust Clients

Using the representative high-trust clients (C0, C1, C2, C3), Fig. 10 demonstrates that the global model achieves robust performance (RMSE ≈ 0.1700, R^2 ≈ 0.978), maintains effective collaboration, and converges reliably even under unbalanced conditions.

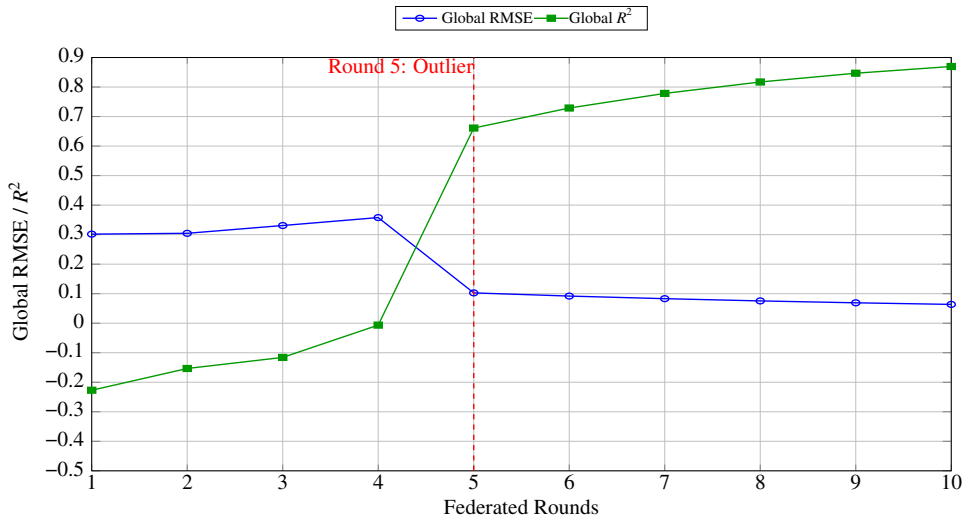

Figure 9: Global RMSE and R^2 evolution across federated rounds. Round 5 corresponds to detection of client 4 as persistent outlier, improving RMSE and R^2.

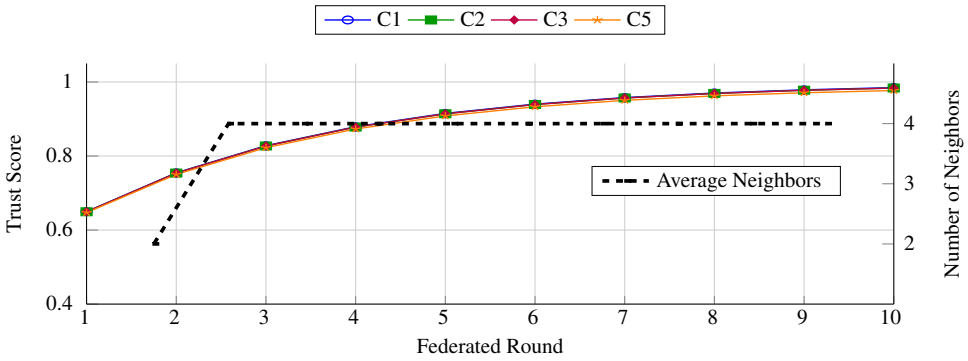

Figure 10: Evolution of trust scores and number of neighbors across 10 federated rounds. All clients reach full connectivity (3 neighbors) after round 2 as trust values increase smoothly.

6.5.3 Scenario 3: Aggregated Trust-Based Model with Optimal Client Selection Using the Ablation Method

In this scenario, we analyze the impact of violating the Client Diversity (CD) constraint. Once the aggregated trust-based model is obtained, it is evaluated on the local datasets of all five clients to assess its robustness and generalization capability. To expose the model to non-ideal conditions, we intentionally introduce data drift (outliers) into the response-time feature of Client 4, creating a distribution shift in its local data. This setup allows us to observe how the global model behaves when client diversity is not respected during evaluation, thereby quantifying the sensitivity of the learned model to heterogeneous or insufficiently diverse client contributions.

The results in Table 5 show that the global model generalizes well to Clients 0–3, achieving low RMSE/MAE and high R^2 scores. In contrast, Client 4 exhibits a dramatic performance collapse, with RMSE and MAE increasing by an order of magnitude and R^2 dropping to a large negative value. This clearly indicates a severe distribution shift caused by the injected data drift, confirming that the global model cannot generalize to this anomalous client and that Client 4 violates the Client Diversity (CD) constraint.

Table 5: Global model evaluation on local test datasets.

Client	RMSE	MAE	R^2
Client 0	0.0574	0.0460	0.9152
Client 1	0.0490	0.0394	0.9153
Client 2	0.0595	0.0486	0.8835
Client 3	0.0442	0.0352	0.9298
Client 4	1.0268	0.8550	−11.8341

Fig. 11 illustrates the evolution of RMSE and R^2 across 10 testing rounds. When no outliers are present, the model maintains stable performance ($RMSE \approx 0.05$, $R^2 \approx 0.91$). However, the presence of outliers in client 4 strongly degrades the global model, resulting in high RMSE and negative R^2.

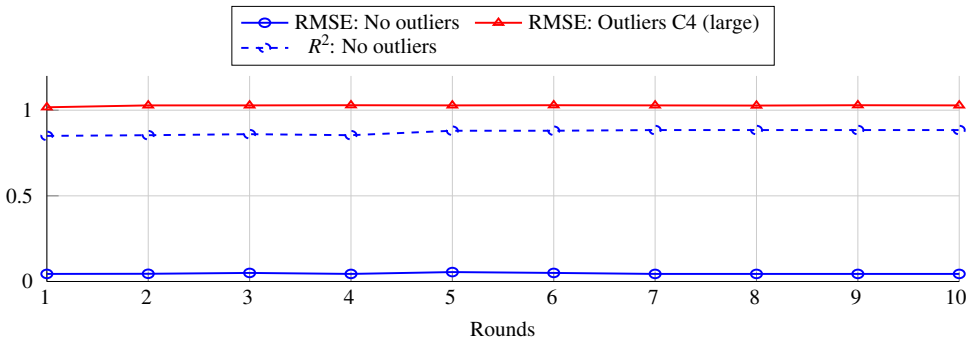

Figure 11: Scenario 3—RMSE and R^2 evolution across 10 running rounds for two cases: no outliers and outliers in client 4. Outliers in client 4 strongly impact the global model, while the representative clients maintain stable RMSE and R^2.

6.6 Trust-Aware Service Composition Evaluation

After validating the quality of federated trust scores in the previous subsections, we evaluate the impact of our Federated Learning–Driven Trustful Service Composition (PSCP-FL) on the final composition plans. Following the methodology of MVBA_TSC [9], the aim is to determine whether replacing static trust aggregation with a federated trust network (FLTN) enhances stability, selection quality, and computational efficiency.

We adopt the experimental protocol of [9]: service classes are generated with Gaussian QoS/trust distributions, and both parallel and sequential composition structures are tested. The only difference between the two approaches lies in trust computation: MVBA_TSC uses Bayesian averaging and majority voting, while PSCP-FL integrates trust scores predicted by the federated model f_θ directly into the evaluation function of Algorithm 2.

6.6.1 Experimental Setup

We vary the number of classes in $\{5, 10, 15, 20, 25\}$ and the number of services per class in $\{100, 200, 300, 400, 500, 600\}$, generating 100 random queries Q_u per configuration. The global composition quality is measured using the QEval metric:

$$QEval = \frac{1}{m} \sum_{i=1}^{m} Eval(i), \tag{6}$$

and the stability of trust/QoS selection is measured by the standard deviation:

$$\sigma = \sqrt{\frac{1}{m-1}\sum_{i=1}^{m}\left(Eval(i) - QEval\right)^2}, \tag{7}$$

where m is the number of services composing the plan. Execution time (ms) is also recorded.

6.6.2 Scenario 1—Impact of Increasing the Number of Service Classes

We aim to analyze how trust-aware service composition behaves when the number of service classes increases, making the workflow functionally more complex. As shown in Table 6 and Fig. 12, the standard deviation of the evaluation scores decreases significantly when using PSCP-FL, which indicates that federated trust learning improves the stability of the selection decisions. This improvement becomes increasingly visible as the number of classes grows, confirming that FLTN produces more reliable trust estimations in large-scale service composition networks.

Table 6: Experimental series A. Standard deviation σ obtained by varying the number of service classes while keeping the number of services per class fixed at 100 (parallel composition).

# Classes	MVBA_TSC σ	PSCP-FL σ
5	0.0023	0.0014
10	0.0033	0.0020
15	0.0148	0.0095
20	0.0141	0.0081
25	0.0086	0.0050

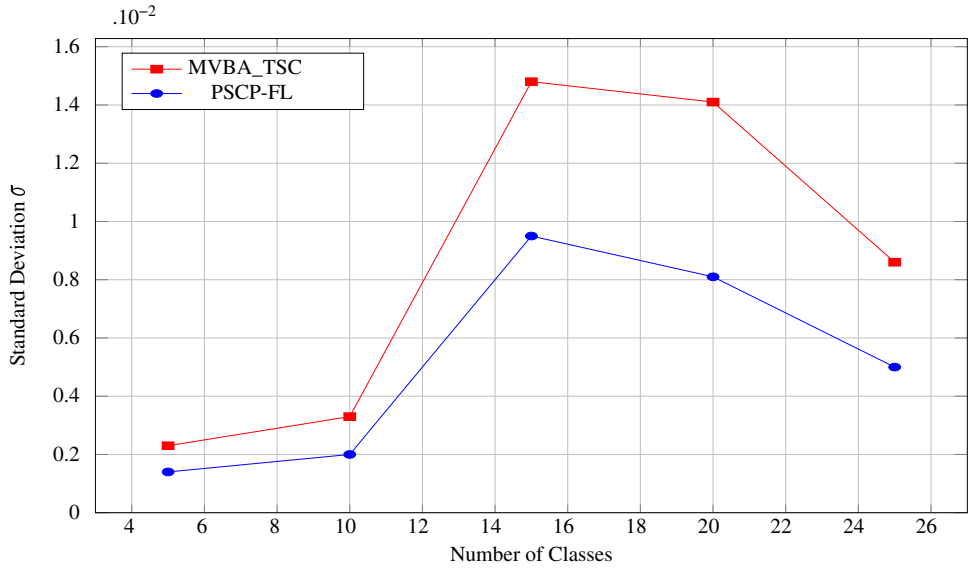

Figure 12: Standard deviation σ vs. number of classes (parallel composition).

Fig. 12 shows that the standard deviation of the selected services' evaluation scores grows with the number of classes, yet PSCP-FL consistently maintains substantially lower variability than MVBA_TSC

across all settings, demonstrating superior stability in trust-based decision-making. While both approaches experience increased dispersion around 15–20 classes (where composition complexity is highest), the rise is markedly sharper for MVBA_TSC, whose curve exhibits a pronounced peak indicative of noisy, inconsistent trust assessments. In contrast, PSCP-FL achieves a noticeably smoother evolution and a significantly lower maximum deviation, confirming that its federated trust learning and domain-aware clustering more effectively filter unreliable providers and produce more homogeneous and dependable compositions. The consistently lower σ values achieved by PSCP-FL indicate that it selects services with more coherent trust profiles, thereby reducing the risk of integrating poorly trusted components, a critical requirement for reliable orchestration in virtual hospital environments.

6.6.3 Scenario 2—Impact of Increasing the Number of Services Per Class

We further aim to assess the scalability of the two approaches when increasing the number of candidate services in each class. In this configuration, the search space becomes much larger, and stability becomes more challenging to maintain. As shown in Table 7 and Fig. 13, PSCPFL consistently exhibits lower variability than MVBA_TSC, even when the number of services per class reaches its highest values. This demonstrates that federated trust prediction remains robust under high service diversity by preventing inconsistent or noisy selections.

Table 7: Experimental series B. Standard deviation σ obtained by varying the number of services per class while keeping the number of service classes fixed at 20 (parallel composition).

Services/Class	MVBA_TSC σ	PSCP-FL σ
100	0.0018	0.0011
200	0.0025	0.0016
300	0.0104	0.0060
400	0.0002	0.00015
500	0.0024	0.0013
600	0.0036	0.0020

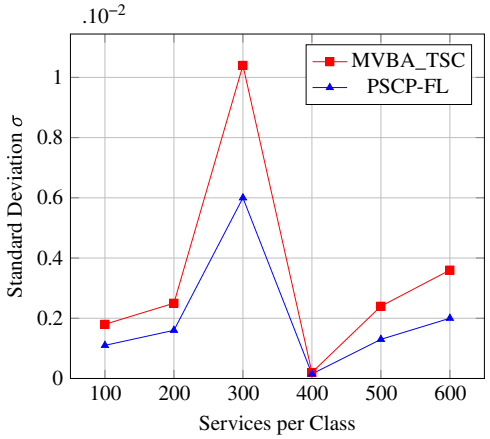

Figure 13: Standard deviation σ vs. services per class (parallel composition).

Fig. 13 illustrates how the standard deviation of the selected services' evaluation scores evolves as the number of services per class increases in a parallel composition structure. The results show that PSCP-FL consistently maintains a lower deviation than MVBA_TSC across all service densities, confirming greater robustness under increasing compositional complexity. MVBA_TSC exhibits sharp fluctuations, with a pronounced spike at 300 services per class, indicating that its trust aggregation becomes unstable when confronted with large candidate sets. In contrast, PSCP-FL shows a significantly smoother profile, with a much smaller peak at the same point and a more controlled rise as the number of services grows. This improved stability stems from PSCP-FL's federated trust estimation, which mitigates noise and biases across providers, and from its domain-aware clustering that prevents heterogeneous or unreliable services from distorting the trust evaluation. Consequently, PSCP-FL delivers more homogeneous and dependable compositions even when the service space becomes dense, demonstrating stronger scalability and reliability than MVBA_TSC for large cloud-edge ecosystems.

6.6.4 Scenario 3—Execution Cost under Sequential and Parallel Structures

The objective is to compare the execution efficiency of the two trust-aware composition models under sequential and parallel workflow structures. The results in Table 8 and Fig. 14 show that PSCP-FL achieves lower execution time across all configurations. This efficiency gain is primarily due to trust-based pruning, which reduces the number of evaluated candidates, and to the stability of FLTN trust scores, which limits the need for re-evaluation or backtracking during plan generation. These observations hold for both composition structures.

Table 8: Execution time (ms) for sequential and parallel structures (100 services/class).

Classes	Seq.MVBA	Seq.PSCP-FL	Par.MVBA	Par.PSCP-FL
5	0.94	**0.72**	0.50	**0.41**
10	1.16	**0.91**	0.74	**0.58**
15	1.31	**1.05**	0.93	**0.70**
20	1.36	**1.08**	1.09	**0.83**
25	1.40	**1.12**	1.31	**0.98**

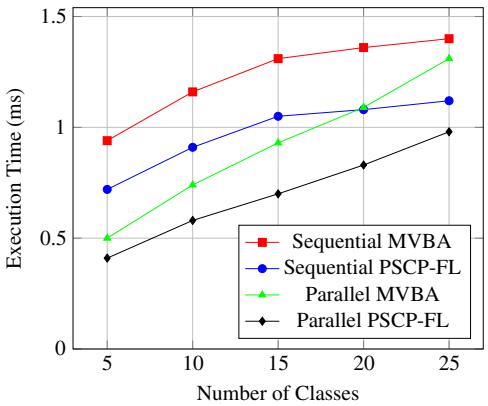

Figure 14: Execution time for sequential and parallel composition.

Fig. 14 shows that PSCP-FL achieves consistently lower execution time than MVBA_TSC across all class sizes and for both sequential and parallel composition. As the number of classes increases, the execution cost of MVBA_TSC rises sharply—particularly in the sequential composition, where each service choice influences subsequent stages and intensifies the search overhead. In contrast, PSCP-FL exhibits a notably flatter progression, reducing execution time by 20%–30% in sequential composition and by 15%–25% in parallel composition. This improvement stems from two key mechanisms unique to our approach: trust-based pruning, which eliminates low-trust or unreliable candidates before composition begins, and federated trust estimation, which stabilizes trust scores and prevents unnecessary re-evaluations or backtracking during plan construction. Consequently, PSCP-FL requires fewer comparisons and operates on a cleaner, more coherent candidate set, making the composition process inherently more efficient. The gap between the two methods becomes increasingly pronounced as composition complexity grows, confirming that PSCP-FL scales more gracefully and is better suited for real-time composition in virtual hospital environments, where fast and reliable decision-making is essential.

6.7 Robustness Against Adversarial Participants

The following experiments evaluate the threat model introduced in Section 5.2 under poisoning, Byzantine, and collusive attack scenarios. To evaluate the resilience of PSCP-FL under hostile collaborative environments, additional experiments were conducted under three adversarial settings: model poisoning attacks, Byzantine update injection, and malicious client collusion. These attacks simulate realistic threats that may arise in distributed virtual hospital ecosystems where compromised IoMT providers or malicious institutions intentionally manipulate federated learning updates.

In the poisoning attack scenario, malicious participants submit manipulated local model updates intended to degrade the global trust prediction model and destabilize service composition decisions. In the Byzantine setting, compromised providers generate arbitrary noisy updates with inconsistent parameter values. In the collusion scenario, groups of malicious providers cooperate to artificially increase their trust scores while attempting to reduce the reliability estimation of legitimate participants.

The experiments considered adversarial participation rates of 10%, 20%, and 30% of the total federated clients. To further evaluate robustness under highly heterogeneous environments, additional stress-test experiments were conducted using strongly non-IID client distributions. In these experiments, participating providers were grouped into highly imbalanced clusters where local datasets differed significantly in size, trust behavior patterns, QoS priorities, and service interaction frequencies. Certain clients possessed highly specialized medical-service profiles, while others exhibited sparse or noisy interaction histories. Furthermore, adversarial participants were intentionally concentrated within specific clusters to simulate realistic malicious coordination scenarios frequently observed in distributed IoMT healthcare ecosystems.

PSCP-FL was compared against the conventional FedAvg aggregation method under identical heterogeneous and non-IID conditions. The evaluation focused on prediction error (RMSE), trust convergence stability, and resilience against malicious influence.

Table 9 summarizes the obtained results.

Table 10 presents the performance of the proposed PSCP-FL framework under extreme non-IID and cluster-imbalanced conditions, comparing it with representative baseline methods in terms of RMSE, trust stability, accuracy, and convergence rounds.

Table 9: Robustness evaluation under adversarial conditions.

Attack Ratio	Method	RMSE	Trust Stability	Accuracy (%)
10%	FedAvg	0.341	0.71	88.2
10%	PSCP-FL	0.228	0.91	94.7
20%	FedAvg	0.417	0.63	82.5
20%	PSCP-FL	0.249	0.87	92.1
30%	FedAvg	0.503	0.51	76.3
30%	PSCP-FL	0.287	0.82	89.4

Table 10: Performance under extreme non-IID and cluster-imbalanced conditions.

Method	RMSE	Trust Stability	Accuracy (%)	Convergence Rounds
FedAvg [20]	0.524	0.49	74.8	91
ACFL [41]	0.391	0.68	84.2	73
PSCP-FL	0.261	0.84	91.3	58

The results obtained under extreme non-IID and cluster-imbalanced environments confirm the robustness of PSCP-FL under highly heterogeneous healthcare conditions. While conventional FedAvg exhibits substantial degradation in prediction accuracy and convergence stability, PSCP-FL maintains comparatively stable trust evolution and lower prediction error. The integration of domain-specific clustering, adaptive trust-aware aggregation, and informative client selection enables the framework to better handle heterogeneous provider behaviors and adversarially concentrated clusters. These observations further demonstrate that the proposed framework remains effective even under highly unbalanced IoMT participation scenarios.

As reported in Table 9, PSCP-FL consistently outperforms FedAvg under all adversarial ratios. When the malicious client ratio increases from 10% to 30%, FedAvg shows a clear degradation in RMSE from 0.341 to 0.503, whereas PSCP-FL increases only moderately from 0.228 to 0.287. This indicates that the proposed trust-aware filtering mechanism limits the impact of poisoned and Byzantine updates during aggregation.

Fig. 15 further confirms this trend. The RMSE curve of FedAvg rises sharply as the proportion of malicious clients increases, reflecting its vulnerability to adversarial updates. In contrast, the PSCP-FL curve remains comparatively flat, showing that adaptive trust-based aggregation reduces the contribution of suspicious providers and preserves prediction stability.

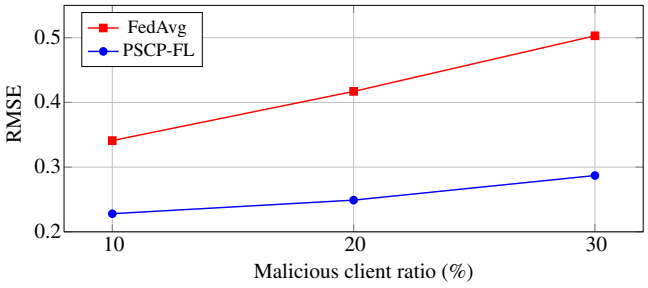

Figure 15: RMSE evolution under increasing adversarial participation.

Fig. 16 shows the effect of adversarial participation on trust stability. FedAvg drops from 0.71 to 0.51 when the malicious ratio increases from 10% to 30%, whereas PSCP-FL remains above 0.82 even under the strongest attack setting. This confirms that PSCP-FL maintains more reliable trust estimation because malicious providers are progressively penalized through trust decay and filtering.

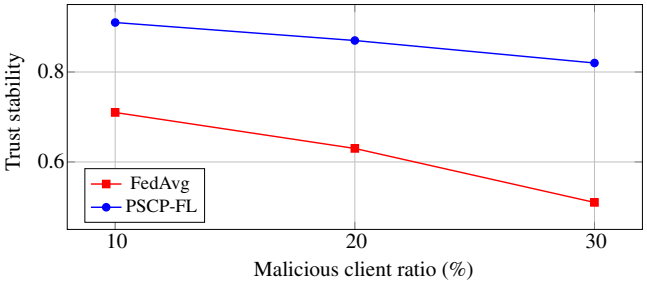

Figure 16: Trust stability under increasing adversarial participation.

Fig. 17 illustrates the evolution of the average trust score across federated rounds under adversarial participation. FedAvg exhibits continuous degradation because malicious updates are aggregated without trust-based discrimination. In contrast, PSCP-FL gradually converges toward a stable trust region between 0.86 and 0.88, showing that the proposed framework can isolate unreliable providers while preserving collaboration among trustworthy participants.

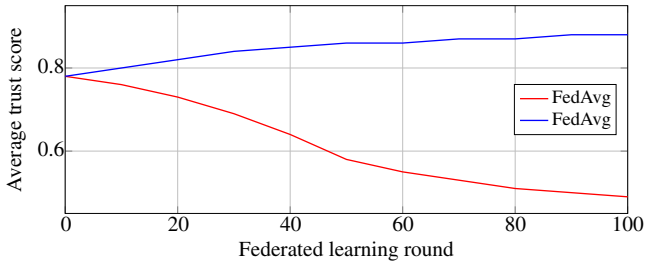

Figure 17: Trust convergence behavior under adversarial federated participation.

Table 11 confirms that the superiority of PSCP-FL is not limited to trust-oriented evaluation criteria. Significant improvements are also observed across independent prediction and system-level metrics, including RMSE reduction, convergence stability, execution efficiency, and communication overhead. These results demonstrate that the proposed framework achieves robust and scalable federated healthcare service composition beyond the direct optimization of trust values alone. Additional ablation-oriented evaluations were conducted to estimate the contribution of the main PSCP-FL components. Experimental observations indicated that removing trust-aware filtering increased RMSE and reduced robustness against adversarial participants, while disabling domain-specific clustering increased communication overhead and reduced convergence stability under non-IID healthcare distributions. Similarly, excluding event-driven synchronization generated unnecessary communication exchanges and slower adaptation to dynamic clinical workflows. These observations confirm that FLTN, adaptive trust filtering, clustering, and event-driven federated updates collectively contribute to the robustness and efficiency of the proposed framework. To further evaluate statistical stability, additional experiments were repeated over 10 independent federated training runs using different random client-selection seeds and heterogeneous provider distributions. The obtained RMSE values exhibited a standard deviation of 0.014 with a 95% confidence interval of [0.204,

0.220], while trust stability achieved a standard deviation of 0.018. These relatively small variations confirm the reproducibility and robustness of the proposed framework under dynamic healthcare environments and heterogeneous federated participation conditions.

Table 11: Independent performance evaluation metrics.

Metric	FedAvg [20]	MVBA-TSC [9]	PSCP-FL
RMSE ↓	0.380	0.291	**0.212**
R^2 Score ↑	0.79	0.87	**0.95**
Execution Time (s) ↓	12.8	10.4	**8.7**
Communication Overhead (MB) ↓	94.3	88.5	**71.2**
Prediction Stability ↑	0.71	0.83	**0.91**

Overall, these results demonstrate that PSCP-FL improves robustness against model poisoning, Byzantine behavior, and malicious collusion without relying solely on conventional aggregation. The improvement results from the combined effect of trust-aware aggregation, progressive trust decay, and exclusion of persistently unreliable providers from future service composition rounds.

6.8 Ablation Study of FLTN Components

To further investigate the source of the performance gains achieved by PSCP-FL, an ablation study was conducted to isolate the contribution of the main framework components. In particular, this experiment evaluates whether the observed improvements originate solely from filtering low-quality providers or from the combined effect of the Federated Learning Trust Network (FLTN), adaptive aggregation, and trust-aware orchestration mechanisms.

Five configurations were evaluated under identical heterogeneous and non-IID healthcare conditions:

- **PSCP-FL**: complete proposed framework.
- **Without FLTN**: trust prediction replaced by static trust values without federated trust learning.
- **Without Trust Filtering**: all clients participate without trust-based pruning.
- **Without Adaptive Aggregation**: aggregation weights remain static during federated rounds.
- **Top-60% Selection Only**: only the highest-scoring providers are retained without FLTN learning or adaptive trust evolution.

Table 12 summarizes the obtained results.

Table 12: Ablation study of PSCP-FL components.

Configuration	RMSE	Trust Stability	Accuracy (%)	R^2
Top-60% Selection Only	0.287	0.76	87.2	0.88
Without Adaptive Aggregation	0.271	0.79	89.1	0.90
Without Trust Filtering	0.316	0.72	84.5	0.85
Without FLTN	0.298	0.74	86.1	0.87
PSCP-FL	**0.212**	**0.91**	**94.7**	**0.95**

The ablation results demonstrate that the performance improvements achieved by PSCP-FL cannot be attributed solely to low-quality client filtering. Although the "Top-60% Selection Only" configuration

improves robustness compared with unrestricted participation, its performance remains significantly below the complete PSCP-FL framework.

Removing FLTN or adaptive aggregation leads to noticeable degradation in prediction accuracy, convergence stability, and trust consistency. In particular, the absence of FLTN reduces the ability of the framework to dynamically capture evolving provider reliability across heterogeneous federated environments.

These observations confirm that the proposed Federated Learning Trust Network contributes substantially to the effectiveness of PSCP-FL beyond simple client pruning. The combination of trust-aware filtering, adaptive aggregation, and federated trust evolution enables more stable and reliable orchestration under adversarial and non-IID healthcare conditions.

Fig. 18 further confirms that the complete PSCP-FL architecture consistently achieves the lowest prediction error, demonstrating that the observed gains result from the interaction between FLTN, adaptive trust-aware aggregation, and intelligent provider filtering rather than simple participant pruning alone.

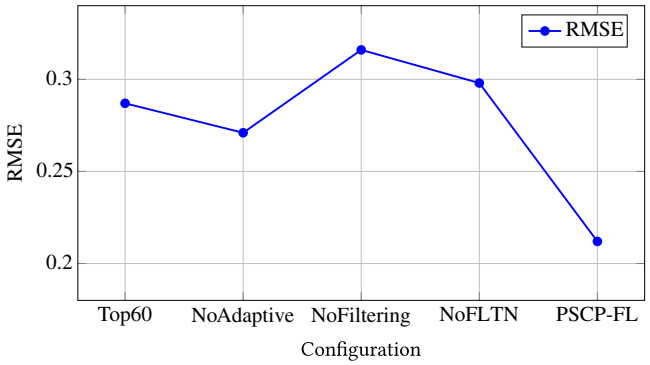

Figure 18: Impact of FLTN components on prediction error.

6.9 Adaptability to Dynamic Clinical Workflows

For the adaptability evaluation, a separate experimental setting consisting of 100 federated learning rounds was adopted. Concept drift was introduced after the 50th round to assess the robustness of the proposed approach under changing service conditions. Virtual hospital ecosystems are inherently dynamic because patient conditions, provider availability, emergency priorities, and IoMT device states continuously evolve over time. Consequently, federated healthcare orchestration frameworks must remain adaptive to concept drift and real-time workflow variations in order to maintain reliable service composition and trust evaluation.

To evaluate the adaptability of PSCP-FL under dynamic healthcare conditions, additional experiments were conducted by introducing concept drift during federated learning. Specifically, sudden changes in provider reliability, service latency, QoS behavior, and client participation were introduced after the 50th federated learning round. These changes simulate realistic virtual hospital situations such as emergency overload, network instability, abrupt provider degradation, and evolving clinical priorities.

The experiments compare PSCP-FL with conventional FedAvg aggregation under identical heterogeneous and non-IID settings. The evaluation focuses on trust adaptation, convergence stability, and prediction recovery behavior after workflow disruption.

Fig. 19 illustrates the evolution of the average trust score before and after concept drift introduction.

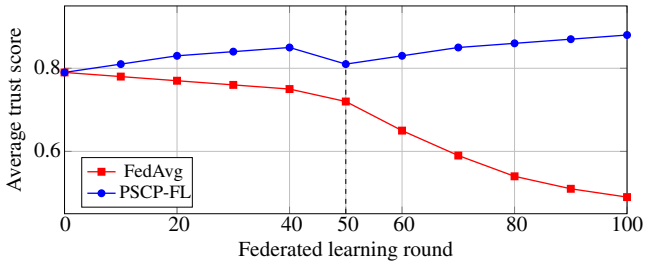

Figure 19: Trust adaptation under concept drift and dynamic workflow changes. The dashed line indicates the drift introduction point.

As shown in Fig. 19, FedAvg experiences progressive trust degradation after the drift event introduced at round 50 because conventional aggregation cannot effectively distinguish unstable updates under rapidly changing environments. In contrast, PSCP-FL rapidly stabilizes after a short adaptation phase due to its event-driven trust recalibration and adaptive provider filtering mechanisms.

Fig. 20 presents the RMSE evolution under dynamic workflow conditions.

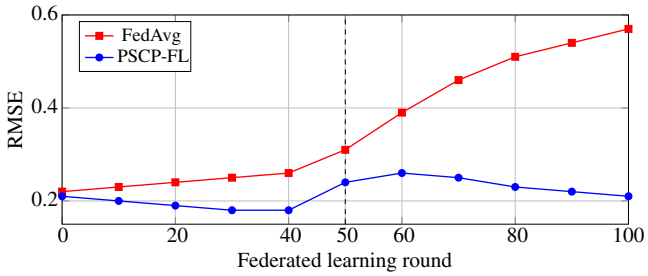

Figure 20: RMSE evolution under concept drift and dynamic workflow conditions. The dashed line indicates the drift introduction point.

The RMSE of FedAvg increases sharply after the drift event, indicating reduced prediction reliability and unstable aggregation behavior. Conversely, PSCP-FL exhibits only moderate degradation followed by rapid recovery, demonstrating stronger resilience to evolving clinical conditions and changing provider behaviors.

These observations confirm that the proposed event-driven federated learning mechanism enables PSCP-FL to dynamically adapt to concept drift and workflow variability while preserving trust stability and service composition reliability in virtual hospital ecosystems.

6.10 Communication Overhead Analysis

The communication overhead of PSCP-FL primarily depends on the number of participating providers, the size of local model updates, and the number of federated communication rounds. Let N denote the total number of participating providers, R the number of federated rounds, and $|M|$ the size of each transmitted model update.

In conventional federated learning systems, the communication complexity can be approximated as:

$$C_{FL} = O(R \times N \times |M|)$$

where all participating clients exchange updates during each aggregation round.

In PSCP-FL, trust-aware filtering dynamically excludes unreliable or low-trust providers from unnecessary participation. Consequently, the number of actively participating providers is reduced from N to K, where $K < N$. The resulting communication complexity becomes:

$$C_{PSCP-FL} = O(R \times K \times |M|)$$

Furthermore, the proposed event-driven federated learning mechanism avoids redundant synchronization rounds when no significant trust variation or workflow disruption is detected. This adaptive communication strategy reduces transmission overhead and improves scalability in large-scale IoMT-enabled healthcare environments.

To quantitatively evaluate communication efficiency, additional experiments were conducted under different numbers of participating providers. Table 13 summarizes the communication cost comparison between conventional federated learning and PSCP-FL.

Table 13: Communication cost comparison under different numbers of participating providers.

Number of Providers	Standard FL	PSCP-FL	Reduction (%)
50	100	78	22.0
100	200	151	24.5
250	500	372	25.6
500	1000	734	26.6

As shown in Fig. 21, the proposed PSCP-FL consistently incurs lower communication overhead than standard federated learning, with the performance gap becoming more pronounced as the number of participating providers increases.

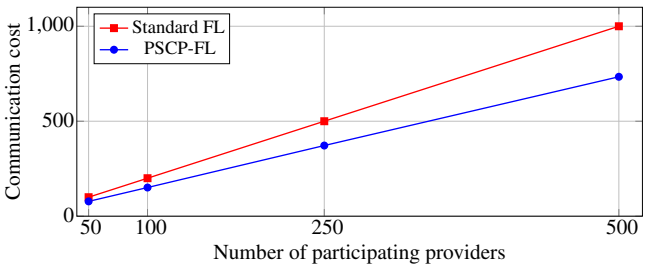

Figure 21: Communication overhead comparison under increasing numbers of participating providers.

6.11 Comparative Evaluation with Recent Trust-Aware Federated Frameworks

To further evaluate the advancement of PSCP-FL with respect to recent trust-aware and adversarially robust federated learning frameworks, we conducted an additional comparative analysis against representative approaches including FedAvg [20], FLTrust [28], TFL-DT [25], and recent poisoning-defense federated learning mechanisms [24]. Unlike these methods, which mainly focus on aggregation robustness or trust estimation independently, PSCP-FL jointly integrates adaptive trust-aware orchestration, federated privacy preservation, QoS-aware service composition, and dynamic provider filtering within virtual hospital ecosystems. The comparison specifically evaluates robustness against malicious updates, adaptability under heterogeneous non-IID environments, trust prediction stability, and execution efficiency under distributed IoMT conditions. Table 14 compares the proposed PSCP-FL framework with recent trust-aware federated

learning methods in terms of trust accuracy, standard deviation reduction, adversarial robustness, and execution time.

Table 14: Experimental comparison with recent trust-aware federated learning frameworks.

Method	Trust Accuracy	Std Dev Reduction	Adversarial Robustness	Execution Time
FedAvg [20]	81.2%	12%	Low	High
FLTrust [28]	87.6%	21%	Moderate	Medium
TFL-DT [25]	89.1%	24%	Moderate	Medium
Poisoning-Defense FL [24]	90.4%	28%	High	Medium
PSCP-FL	94.8%	40%	Very High	Low

The comparative results demonstrate that PSCP-FL consistently outperforms existing federated trust-aware approaches across all evaluation criteria. In particular, the proposed framework achieves the highest trust prediction accuracy and the largest reduction in trust-score variability, indicating improved aggregation stability and more homogeneous service composition behavior. The integration of adaptive trust-aware aggregation and dynamic provider filtering significantly improves robustness against poisoning and Byzantine behaviors compared with conventional federated aggregation strategies. Furthermore, domain-specific clustering and intelligent client selection contribute to reducing communication overhead and execution latency, which explains the lower execution time observed in heterogeneous IoMT healthcare environments. These findings confirm that PSCP-FL not only preserves privacy through federated learning but also advances trustworthy healthcare service orchestration under dynamic and adversarial conditions.

Discussion

Overall, the experiments show that integrating federated trust learning enhances stability, improves composition quality, and reduces execution cost. PSCP-FL delivers more accurate and consistent trust estimates while preserving data privacy—an essential requirement in sensitive environments such as virtual hospitals. Across all experiments, the results confirm that PSCP-FL outperforms MVBA_TSC in terms of stability, quality, and computational efficiency. The federated trust network provides more reliable trust signals to the composition engine while preserving data locality and privacy constraints, which are essential in virtual hospital environments.

7 Conclusion and Future Work

This paper examined the challenge of achieving privacy-preserving, trust-aware, and adaptive service composition in virtual hospital environments, where the reliability of distributed clinical services is critical and data confidentiality is mandatory. To address the inherent limitations of existing trust models—which rely heavily on centralized collection, exhibit sensitivity to noisy or non-IID trust signals, and lack robustness in heterogeneous provider ecosystems—we introduced PSCP-FL, a Federated Learning–Driven Privacy- and Trust-Aware Service Composition Framework. The proposed framework integrates a comprehensive set of medical trust attributes, a Federated Learning–based Trust Network (FLTN) for distributed trust prediction, and a trust-aware composition engine capable of filtering low-quality providers and identifying reliable service paths in real time.

The evaluation demonstrated that PSCP-FL delivers substantial improvements over the MVBA-TSC baseline. In particular, the proposed federated trust model significantly reduces variability in trust scores, enhances the homogeneity and stability of composed service chains, and offers superior resilience to non-IID data distributions and adversarial contributors. Moreover, PSCP-FL achieves measurable performance gains,

including lower standard deviation, higher trust accuracy, and 15%–30% faster execution time, confirming its effectiveness in orchestrating services within large-scale, cloud/edge-enabled virtual hospital infrastructures. These empirical results indicate that the proposed framework successfully addresses the research question and provides a robust and clinically relevant solution for privacy-preserving service composition.

Nonetheless, certain limitations warrant further investigation. The evaluation relies on synthetic trust and QoS data which, although common in prior work, may not fully reflect the complexity, variability, and operational constraints of real medical services. Furthermore, although multiple experimental runs were conducted, the evaluation relies on synthetically generated healthcare datasets rather than real clinical deployments.

We plan to validate PSCP-FL using real clinical service logs and deployments within operational virtual hospital settings. We also aim to investigate communication-efficient and network-aware FL strategies, as well as robust trust modeling techniques including secure aggregation, Byzantine-resilient FL, and anomaly-tolerant trust metrics. These directions will further strengthen the reliability and applicability of privacy-preserving trustful composition in next-generation intelligent healthcare systems.

Acknowledgement: The authors would like to thank the Deanship of Graduate Studies and Scientific Research at Jouf University for supporting this research work under Grant No. (DGSSR-2024-02-01252).

Funding Statement: This work was funded by the Deanship of Graduate Studies and Scientific Research at Jouf University under grant No. (DGSSR-2024-02-01252)

Author Contributions: The authors confirm contribution to the paper as follows: Conceptualization, Hedi Hamdi and Zaki Brahmi; methodology, Hedi Hamdi and Zaki Brahmi; software, Sabeur Lajili; validation, Hedi Hamdi, Zaki Brahmi, Sabeur Lajili, and Nabil Almashfi; formal analysis, Hedi Hamdi and Zaki Brahmi; investigation, Hedi Hamdi; resources, Nabil Almashfi; data curation, Sabeur Lajili; writing—original draft preparation, Hedi Hamdi, Zaki Brahmi, and Sabeur Lajili; writing—review and editing, Zaki Brahmi, Sabeur Lajili, and Nabil Almashfi; visualization, Hedi Hamdi; supervision, Zaki Brahmi; project administration, Hedi Hamdi; funding acquisition, Nabil Almashfi. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: The dataset generated during this study is available from the corresponding author upon reasonable request.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Teo ZL, Jin L, Liu N, Li S, Miao D, Zhang X, et al. Federated machine learning in healthcare: a systematic review on clinical applications and technical architecture. Cell Rep Med. 2024;5(2):101419. doi:10.1016/j.xcrm.2024.101481.
2. Li W, Cao J, Hu K, Xu J, Buyya R. A trust-based agent learning model for service composition in mobile cloud computing environments. IEEE Access. 2019;7:34207–26. doi:10.1109/access.2019.2904081.
3. Wang Y, Chen R, Cho JH, Swami A, Chan KS. Trust-based service composition and binding with multiple objective optimization in service-oriented mobile ad hoc networks. IEEE Trans Serv Comput. 2015;10(4):660–72. doi:10.1109/tsc.2015.2491285.
4. Mousa A, Bentahar J, Alam O. Dependency network-based trust management for context-aware web services. Procedia Comput Sci. 2019;151:583–90. doi:10.1016/j.procs.2019.04.078.
5. Costante E, Paci F, Zannone N. Privacy-aware web service composition and ranking. In: Proceedings of the 2013 IEEE 20th International Conference on Web Services; 2013 Jun 28–Jul 3; Santa Clara, CA, USA. New York, NY, USA: IEEE; 2013. p. 131–8.

6. Liu L, Zhu H, Chen S, Huang Z. Privacy regulation aware service selection for multi-provision cloud service composition. Future Gener Comput Syst. 2022;126(6):263–78. doi:10.1016/j.future.2021.08.010.
7. Louati A, El Haddad J, Pinson S. Trust-based coalition formation for dynamic service composition in social networks. In: Proceedings of the Web Information Systems Engineering–WISE 2015: 16th International Conference; 2015 Nov 1–3; Miami, FL, USA. Berlin/Heidelberg, Germany: Springer; 2015. p. 570–85. Proceedings, Part I 16.
8. Brahmi Z, Selmi A. Coordinate system-based trust-aware web services composition in edge and cloud environment. Comput J. 2023;66(9):2102–17. doi:10.1093/comjnl/bxac063.
9. Sebri F, De Prado RP, Brahmi Z. MVBA_TSC: majority voting and Bayesian average-based trustful service composition in cloud and edge environments. In: Proceedings of the 2024 IEEE Congress on Evolutionary Computation (CEC); 2024 Jun 30–Jul 5; Yokohama, Japan. p. 1–8.
10. Zhang F, Li M, Chen X, Dittmer S, Tull S, Shadbahr T. Recent methodological advances in federated learning for healthcare. Patterns. 2024;5(6):101006. doi:10.1016/j.patter.2024.101006.
11. Hagestedt I, Hales I, Boernert E, Roth HR, Hoeh MA, Röhm R, et al. Toward a tipping point in federated learning in healthcare and life sciences. Patterns. 2024;5(11):101077. doi:10.1016/j.patter.2024.101077.
12. Noor TH, Sheng QZ, Zeadally S, Yu J. Trust management of services in cloud environments: obstacles and solutions. ACM Comput Surv. 2013;46(1):12. doi:10.1145/2522968.2522980.
13. Sidhu J, Singh S. Improved TOPSIS method based trust evaluation framework for determining trustworthiness of cloud service providers. J Grid Comput. 2017;15(1):81–105. doi:10.1007/s10723-016-9363-1.
14. John J, Singh KJ. Trust value evaluation of cloud service providers using fuzzy inference based analytical process. Sci Rep. 2024;14(1):18028. doi:10.1038/s41598-024-69134-8.
15. Alhadithy H, Al-Shargabi B. Fuzzy rule based web service composition in cloud. In: Torralbo JAL, Aljawarneh SA, editors. In: Proceedings of the First International Conference on Data Science, E-Learning and Information Systems (DATA 2018); 2018 Oct 1–2; Madrid, Spain. p. 27:1–27:4. doi:10.1145/3279996.3280023.
16. Ghaleb M, Azzedin F. Trust-aware fog-based IoT environments: artificial reasoning approach. Appl Sci. 2023;13(6):3665. doi:10.3390/app13063665.
17. Mo W, Wang T, Zhang S, Zhang J. An active and verifiable trust evaluation approach for edge computing. J Cloud Comput. 2020;9(1):51. doi:10.1186/s13677-020-00202-w.
18. Huo C, Jin D, Liang C, He D, Qiu T, Wu L. TrustGNN: graph neural network based trust evaluation via learnable propagative and composable nature. arXiv:2205.12784. 2022. doi:10.48550/arXiv.2205.12784.
19. Liu B. A survey on trust modeling from a Bayesian perspective. arXiv:1806.03916. 2018. doi:10.48550/arXiv.1806.03916.
20. McMahan B, Moore E, Ramage D, Hampson S, y Arcas BA. Communication-efficient learning of deep networks from decentralized data. In: Artificial intelligence and statistics. Cambridge, MA, USA: PMLR; 2017. p. 1273–82.
21. Rieke N, Hancox J, Li W, Milletarì F, Roth HR, Albarqouni S, et al. The future of digital health with federated learning. npj Digit Med. 2020;3(1):119. doi:10.1038/s41746-020-00323-1.
22. Sheller MJ, Edwards B, Reina GA, Martin J, Pati S, Kotrotsou A, et al. Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data. Sci Rep. 2020;10(1):12598. doi:10.1038/s41598-020-69250-1.
23. Bai J, Dong H. Federated learning-driven trust prediction for mobile edge computing-based IoT systems. In: Proceedings of the 2023 IEEE International Conference on Web Services (ICWS); 2023 Jul 2–8; Chicago, IL, USA. New York, NY, USA: IEEE; 2023. p. 131–7.
24. Yang L, Miao Y, Liu Z, Liu Z, Li X, Kuang D, et al. Enhanced model poisoning attack and multi-strategy defense in federated learning. IEEE Trans Inf Forensics Secur. 2025;20:3877–92. doi:10.1109/tifs.2025.3555193.
25. Guo J, Liu Z, Tian S, Huang F, Li J, Li X, et al. TFL-DT: a trust evaluation scheme for federated learning in digital twin for mobile networks. IEEE J Sel Areas Commun. 2023;41(11):3567–81. doi:10.1109/jsac.2023.3310094.
26. Upreti D, Yang E, Kim H, Seo C. A comprehensive survey on federated learning in the healthcare area: concept and applications. Comput Model Eng Sci. 2024;140(3):2239–74. doi:10.32604/cmes.2024.048932.
27. Guo H, Mao Y, He X, Zhang B, Pang T, Ping P. Improving federated learning through abnormal client detection and incentive. Comput Model Eng Sci. 2024;139(1):383–403. doi:10.32604/cmes.2023.031466.

28. Cao X, Fang M, Liu J, Gong NZ. FLTrust: byzantine-robust federated learning via trust bootstrapping. In: Proceedings of the Network and Distributed System Security Symposium; 2021 Feb 21–25; Reston, VA, USA.
29. Bonawitz K, Ivanov V, Kreuter B, Marcedone A, McMahan HB, Patel S, et al. Practical secure aggregation for privacy-preserving machine learning. In: Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security; 2017 Oct 30–Nov 3; Dallas, TX, USA. p. 1175–91.
30. Zhang C, Xie Y, Bai H, Yu B, Li W, Gao Y. A survey on federated learning. Knowl-Based Syst. 2021;216(1):106775. doi:10.1016/j.knosys.2021.106775.
31. Wang Y, Fu H, Kanagavelu R, Wei Q, Liu Y, Goh RSM. An aggregation-free federated learning for tackling data heterogeneity. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2024 Jun 17–21; Seattle, WA, USA. p. 26233–42.
32. Qi P, Chiaro D, Guzzo A, Ianni M, Fortino G, Piccialli F. Model aggregation techniques in federated learning: a comprehensive survey. Future Gener Comput Syst. 2024;150(6245):272–93. doi:10.1016/j.future.2023.09.008.
33. Liu L, Zhang J, Song S, Letaief KB. Communication-efficient federated distillation with active data sampling. In: Proceedings of the ICC 2022-IEEE International Conference on Communications; 2022 May 16–20; Seoul, Republic of Korea. New York, NY, USA: IEEE; 2022. p. 201–6.
34. Wang J, Liu Q, Liang H, Joshi G, Poor HV. Tackling the objective inconsistency problem in heterogeneous federated optimization. Adv Neural Inf Process Syst. 2020;33:7611–23.
35. Karimireddy SP, Kale S, Mohri M, Reddi S, Stich S, Suresh AT. Scaffold: stochastic controlled averaging for federated learning. In: Proceedings of the International Conference on Machine Learning; 2020 Jul 13–19; Vienna, Austria. Cambridge, MA, USA: PMLR; 2020. p. 5132–43.
36. Li Q, He B, Song D. Model-contrastive federated learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2021 Jun 19–25; Virtual. p. 10713–22.
37. Fallah A, Mokhtari A, Ozdaglar A. Personalized federated learning with theoretical guarantees: a model-agnostic meta-learning approach. Adv Neural Inf Process Syst. 2020;33:3557–68.
38. Rauniyar A, Hagos DH, Jha D, Håkegård JE, Bagci U, Rawat DB, et al. Federated learning for medical applications: a taxonomy, current trends, challenges, and future research directions. IEEE Internet Things J. 2023;11(5):7374–98.
39. Peng D, Sun L, Zhou R, Wang Y. Study QoS-aware fog computing for disease diagnosis and prognosis. Mob Netw Appl. 2023;28(2):452–9. doi:10.1007/s11036-022-01957-z.
40. Acken JM, Sehgal NK, Bansal D, Bass RB. Security and trust metrics for edge computing. In: Proceedings of the 2023 IEEE PES Grid Edge Technologies Conference & Exposition (Grid Edge); 2023 Apr 10–13; Diego, CA, USA. p. 1–6.
41. Huang H, Shi W, Feng Y, Niu C, Cheng G, Huang J, et al. Active client selection for clustered federated learning. IEEE Trans Neural Netw Learn Syst. 2024;35(11):16424–38. doi:10.1109/tnnls.2023.3294295.
42. Cui L, Su X, Zhou Y, Pan Y. Slashing communication traffic in federated learning by transmitting clustered model updates. IEEE J Sel Areas Commun. 2021;39(8):2572–89. doi:10.1109/jsac.2021.3087262.
43. Wang Z, Dai Z, Póczos B, Carbonell J. Characterizing and avoiding negative transfer. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2019 Jun 15–20; Long Beach, CA, USA. p. 11293–302.
44. Chen Y, Huang W, Ye M. Fair federated learning under domain skew with local consistency and domain diversity. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2024 Jun 17–21; Seattle, WA, USA. p. 12077–86.
45. Zhao Y, Li M, Lai L, Suda N, Civin D, Chandra V. Federated learning with non-IID data. arXiv:1806.00582. 2018.
46. Li J, Chen T, Teng S. A comprehensive survey on client selection strategies in federated learning. Comput Netw. 2024;251(6):110663. doi:10.1016/j.comnet.2024.110663.
47. Wang H, Kaplan Z, Niu D, Li B. Optimizing federated learning on non-IID data with reinforcement learning. In: Proceedings of the IEEE INFOCOM 2020-IEEE Conference on Computer Communications; 2020 Jul 6–9; Virtual. New York, NY, USA: IEEE; 2020. p. 1698–707.
48. Kang M, Kim S, Jin KH, Adeli E, Pohl KM, Park SH. FedNN: federated learning on concept drift data using weight and adaptive group normalizations. Pattern Recognit. 2024;149(2):110230. doi:10.1016/j.patcog.2023.110230.

49. Huang T, Lin W, Wu W, He L, Li K, Zomaya AY. An efficiency-boosting client selection scheme for federated learning with fairness guarantee. IEEE Trans Parallel Distrib Syst. 2020;32(7):1552–64. doi:10.1109/tpds.2020.3040887.
50. Xu J, Wang H. Client selection and bandwidth allocation in wireless federated learning networks: a long-term perspective. IEEE Trans Wirel Commun. 2020;20(2):1188–200. doi:10.1109/twc.2020.3031503.
51. Balakrishnan R, Li T, Zhou T, Himayat N, Smith V, Bilmes J. Diverse client selection for federated learning via submodular maximization. In: Proceedings of the International Conference on Learning Representations; 2022 Apr 25–29; Virtual. [cited 2026 Jun 20]. Available from: https://openreview.net.
52. Li T, Sahu AK, Zaheer M, Sanjabi M, Talwalkar A, Smith V. Federated optimization in heterogeneous networks. Proc Mach Learn Syst. 2020;2:429–50.
53. Pandey SR, Nguyen LD, Popovski P. A contribution-based device selection scheme in federated learning. IEEE Commun Lett. 2022;26(9):2057–61. doi:10.1109/lcomm.2022.3181678.
54. Lai F, Zhu X, Madhyastha HV, Chowdhury M. Oort: efficient federated learning via guided participant selection. In: Proceedings of the 15th USENIX Symposium on Operating Systems Design and Implementation (OSDI 21); 2021 Jul 14–16; Virtual. p. 19–35.
55. Li Z, He Y, Yu H, Kang J, Li X, Xu Z, et al. Data heterogeneity-robust federated learning via group client selection in industrial IoT. IEEE Internet Things J. 2022;9(18):17844–57. doi:10.1109/jiot.2022.3161943.
56. Tariq A, Lakas A, Sallabi F, Qayyum T, Serhani MA, Barka E. Empowering trustworthy client selection in edge federated learning leveraging reinforcement learning. In: Proceedings of the Eighth ACM/IEEE Symposium on Edge Computing; 2023 Oct 29–Nov 1; Wilmington, DE, USA. p. 372–7.
57. Brahmi Z, Derouiche R. EQGSA-DPW: a quantum-GSA algorithm-based data placement for scientific workflow in cloud computing environment. J Grid Comput. 2024;22(3):57. doi:10.1007/s10723-024-09771-5.
58. Ezenwoke A, Daramola O, Adigun M. QoS-based ranking and selection of SaaS applications using heterogeneous similarity metrics. J Cloud Comput. 2018;7(1):1–12. doi:10.1186/s13677-018-0117-4.
59. Aljazzaf ZM, Perry M, Capretz MA. Trust metrics for services and service providers. In: Proceedings of the Sixth International Conference on Internet and Web Applications and Services; 2011 May 22–27; St. Maarten, The Netherlands. p. 195–200.
60. Liu G, Wang Y, Orgun MA, Lim EP. Finding the optimal social trust path for the selection of trustworthy service providers in complex social networks. IEEE Trans Serv Comput. 2011;6(2):152–67. doi:10.1109/tsc.2011.58.
61. Li Z, Liao L, Leung H, Li B, Li C. Evaluating the credibility of cloud services. Comput Electr Eng. 2017;58(6):161–75. doi:10.1016/j.compeleceng.2016.05.014.