

Privacy-Preserving Collaborative Intelligence Using Personalized Federated Learning for Heterogeneous Smart Devices

Dr. Daniel Morgan¹, Dr. David Wills²

¹Professor, Department of Artificial Intelligence, Northbridge Institute of Intelligent Computing, USA.

²Assistant Professor, Department of Computer Engineering, Canadian Institute of Advanced Computing, Canada.

Email : d.morgan@gmail.com

Article Received 08-04-2025, Revised 21-04-2026, Accepted 01-06-2026

Author(s) Retains the Copyrights of This Article

Abstract

The increasing use of smart phones, wearable devices, home appliances, industrial sensors, and embedded intelligent systems has created a large and continuously growing source of distributed data. Although these devices can provide valuable information for machine learning applications, transferring their raw data to centralized servers raises concerns regarding privacy, communication overhead, computational cost, and data ownership. Federated learning provides an alternative by allowing devices to train models locally and share model-related information rather than their original datasets. Nevertheless, conventional federated learning can perform poorly when participating devices differ substantially in hardware capabilities, data distributions, network conditions, and application requirements. This research proposes a privacy-preserving collaborative intelligence framework based on personalized federated learning for heterogeneous smart devices. The framework separates common knowledge from device-specific knowledge through a personalized learning mechanism, enabling participating devices to benefit from collaborative training while retaining adaptation to their local characteristics. A resource-aware client selection strategy considers processing capability, battery condition, communication quality, and local dataset characteristics before initiating training rounds. To strengthen privacy protection, the framework incorporates secure aggregation and controlled update protection so that individual model contributions are not directly exposed to the coordinating server. The proposed architecture also accommodates non-identically distributed data by allowing local personalization layers to adapt the shared representation to individual device environments. This approach aims to establish a balance among privacy protection, personalization, communication efficiency, computational requirements, and predictive performance. The framework is applicable to heterogeneous smart-device environments such as intelligent homes, wearable computing, connected vehicles, industrial monitoring, and mobile health applications. The study provides a unified architectural basis for collaborative intelligence in which devices can learn collectively without requiring centralized access to their raw private data.

Keywords

Personalized Federated Learning, Privacy-Preserving Machine Learning, Heterogeneous Smart Devices, Collaborative Intelligence, Secure Aggregation, Distributed Learning, Non-IID Data, Edge Intelligence, Resource-Aware Learning, Smart IoT Devices, On-Device Learning, Data Privacy.

1. Introduction

The rapid expansion of smart phones, wearable devices, connected home appliances, industrial controllers, intelligent transportation systems, and embedded sensing platforms has created an increasingly distributed environment for machine learning. These devices continuously generate data

associated with user behavior, physical environments, operational conditions, and application-specific activities. Conventional machine learning systems generally transfer collected information to centralized servers for storage, processing, and model development. Although centralized processing can simplify model management and provide access to

large computational resources, it also creates challenges related to data ownership, communication requirements, privacy exposure, and dependence on reliable network connectivity. Federated learning has emerged as an alternative distributed learning paradigm in which participating devices retain their local training data and contribute to collaborative model development through model updates rather than transferring their complete datasets to a central location. The original Federated Averaging approach demonstrated that decentralized clients could collaboratively train a shared model while keeping training examples on local devices, establishing an important foundation for privacy-aware distributed machine learning. [1]

Privacy protection in federated learning cannot, however, be reduced simply to keeping raw data on the device. Model updates transmitted during collaborative training may contain information that can potentially reveal characteristics of the underlying training data. Differentially private learning has therefore become an important complementary mechanism for reducing the possibility of extracting sensitive information from model training processes. McMahan et al. investigated differential privacy for recurrent language models in a federated setting and demonstrated the importance of controlling information leakage while maintaining useful model performance. Abadi et al. previously established a practical framework for incorporating differential privacy into deep learning through techniques that constrain individual contributions and introduce controlled randomization during training. These developments indicate that privacy-preserving collaborative intelligence requires consideration of both data locality and the information contained within learning updates. [2], [3]

Secure aggregation provides another important layer of protection by preventing the coordinating server from directly inspecting individual client updates. Instead of allowing the server to observe every model contribution separately, secure aggregation enables the system to recover an aggregate update while protecting the individual contributions of participating devices. Bonawitz et al. introduced a practical secure aggregation protocol designed specifically for privacy-preserving machine learning and demonstrated how cryptographic mechanisms can be incorporated into federated training without requiring the central server to access individual model updates. This property is particularly important for smart-device environments because individual devices may contain highly sensitive information even when their

raw datasets never leave the device. Modern federated systems can consequently combine decentralized data storage, secure aggregation, and additional privacy mechanisms to establish multiple layers of protection. [4]

Another fundamental challenge concerns the communication cost associated with repeated federated training. Smart devices frequently operate through wireless networks with limited bandwidth, intermittent connectivity, constrained data plans, and variable transmission quality. A conventional federated training process may require frequent exchange of large model updates, which can become expensive when the number of participating devices increases. Konečný et al. examined communication-efficient strategies for federated learning and emphasized the importance of reducing the amount of information transmitted between clients and coordinating servers. Their work established communication efficiency as a central design consideration for practical federated systems rather than merely an implementation-level optimization. [5] The computational characteristics of smart devices introduce a second form of heterogeneity. Smartphones, wearable devices, household controllers, industrial sensors, and embedded processors may differ considerably in processor capacity, memory availability, battery condition, storage capacity, and network connectivity. Konečný et al. described federated optimization as an approach for enabling distributed on-device intelligence under such decentralized conditions. The resulting learning process must account for the fact that clients cannot always perform the same amount of computation within the same period. A training architecture designed for heterogeneous devices therefore requires mechanisms for handling different processing capabilities and participation patterns while preserving the collaborative nature of federated learning. [6]

Data heterogeneity represents an equally significant challenge. In realistic smart-device environments, participating clients rarely possess independent and identically distributed datasets. A wearable device may observe data associated with one person's behavior, while a smart-home device may observe environmental patterns specific to one household and an industrial sensor may capture operational conditions associated with a particular machine. Zhao et al. demonstrated that non-IID data can significantly affect the behavior of federated optimization and can cause the collaboratively trained model to perform differently from conventional centralized training. Consequently, simply increasing the number of

participating clients does not necessarily resolve distributional differences among them. Effective collaborative intelligence must explicitly account for differences in local data distributions. [7]

Li et al. subsequently presented a comprehensive perspective on federated learning challenges and identified statistical heterogeneity, system heterogeneity, communication constraints, privacy, and security as interconnected concerns. These challenges are particularly relevant when federated learning is deployed across heterogeneous smart devices rather than homogeneous organizational servers. A system may contain devices with different amounts of local data, different computational resources, different communication conditions, and different application objectives. Such variation makes a single global model insufficient for many practical applications because the model that performs well across the overall population may not provide equally effective behavior for every individual device. [8]

Heterogeneous network conditions further complicate collaborative training. Li et al. studied federated optimization in heterogeneous networks and highlighted the difficulties created by differences in client resources, local objectives, communication conditions, and participation behavior. In a smart-device environment, some clients may possess stable high-speed connections and sufficient battery capacity, whereas others may operate under intermittent connectivity or severe energy limitations. A federated framework that treats every device identically may consequently create synchronization delays, inefficient resource utilization, and unnecessary communication. Resource-aware participation and adaptive learning strategies are therefore important for maintaining practical training efficiency. [9]

Personalization provides a promising direction for addressing the limitations of a single global model. Instead of forcing every participating device to use exactly the same final model, personalized federated learning allows a shared component to capture common knowledge while maintaining device-specific information that reflects local characteristics. Li et al. introduced Ditto as a personalized federated learning approach designed to provide locally adapted models while retaining the benefits of collaborative training. The principle is particularly suitable for smart devices because common patterns can be learned collectively while device-specific variations can remain locally represented. This approach creates a more flexible relationship between global knowledge

and local adaptation than conventional one-model-for-all federated learning. [10]

The broader development of personalized federated learning has produced multiple strategies for addressing client heterogeneity. Tan et al. reviewed personalized federated learning methods and highlighted approaches that modify model structures, aggregation procedures, optimization processes, or local adaptation mechanisms according to client characteristics. Personalization is therefore not limited to adding a small local layer to a global model; it represents a broader design principle in which collaborative learning is combined with client-specific adaptation. For heterogeneous smart devices, this distinction is important because differences in sensing environments, data distributions, hardware capabilities, and application requirements may exist simultaneously. [11]

FedBN provides another perspective on heterogeneous federated learning by retaining batch-normalization statistics locally rather than forcing them to be globally aggregated. Li et al. developed FedBN specifically for federated learning with non-IID feature distributions, demonstrating that preserving certain local model characteristics can improve learning under feature heterogeneity. This principle supports the central motivation of the present research: a collaborative model should share information that is broadly useful while retaining local components that capture characteristics specific to individual devices. The proposed framework extends this concept toward a broader privacy-preserving architecture in which personalization, secure aggregation, resource-aware participation, and heterogeneous device characteristics are considered together. [12]

The preceding developments demonstrate that federated learning has progressed from a basic decentralized averaging mechanism toward a broader collaborative intelligence paradigm involving privacy protection, communication efficiency, personalization, and heterogeneous resource management. Nevertheless, several challenges remain when these requirements occur simultaneously. A practical smart-device environment cannot assume identical hardware, stable connectivity, balanced datasets, or homogeneous user requirements. Furthermore, keeping raw data local does not by itself guarantee complete privacy, while secure aggregation alone does not address model personalization or resource constraints. The research presented in this study therefore focuses on an integrated privacy-preserving personalized federated learning framework designed specifically for heterogeneous smart devices.

The proposed framework combines a shared collaborative representation with device-specific personalization, secure aggregation, controlled update protection, and resource-aware client selection to establish a collaborative learning environment in which devices can benefit from collective knowledge while retaining local adaptation and stronger control over sensitive information.

2. Related Work

Bonawitz et al. investigated practical secure aggregation for privacy-preserving machine learning and established an important mechanism for preventing a coordinating server from directly observing individual client updates. Their approach enables a server to obtain an aggregate result without exposing each participant's contribution, making secure aggregation particularly relevant to federated learning environments containing sensitive user or device information. The work also demonstrates that privacy protection must extend beyond the physical location of training data because model updates themselves may become privacy-sensitive objects during collaborative optimization. This observation forms an important basis for the privacy layer of the proposed framework. [4]

Zhao et al. examined federated learning under non-IID data and showed that differences in local data distributions can substantially affect collaborative model training. In real smart-device environments, the distribution of observations can vary naturally from one device to another because users, locations, operating conditions, and sensing modalities are different. A global model trained by directly combining these heterogeneous contributions may therefore fail to represent important device-specific characteristics. Their findings provide the motivation for incorporating personalization into the proposed architecture rather than relying exclusively on conventional global aggregation. [7]

Li et al. surveyed federated learning challenges, methods, and future directions and identified data heterogeneity, system heterogeneity, communication efficiency, privacy, and security as major issues influencing practical deployment. Their analysis demonstrates that federated learning should be considered as a distributed systems problem as well as a machine learning problem. In heterogeneous smart-device environments, these dimensions interact directly: a device with limited energy may also have limited communication capacity, while a device with abundant computational resources may observe a data distribution substantially different from other clients.

These observations support the need for a framework that jointly considers learning quality and device conditions. [8]

Li et al. investigated federated optimization under heterogeneous network conditions and emphasized that differences among participating clients can create optimization and systems-level difficulties. Devices may participate at different speeds, possess unequal computational resources, or experience different communication conditions. Such differences can affect synchronization and overall training efficiency. Their work is directly relevant to resource-aware federated learning because it demonstrates why a uniform training strategy may not be appropriate for a highly diverse client population. The proposed framework addresses this issue through client selection based on resource and communication characteristics. [9]

Li et al. proposed Ditto as a personalized and robust federated learning approach in which local models can adapt to individual client characteristics while remaining connected to a shared learning process. Personalization is especially important when clients have substantially different data distributions because a single global model may represent common patterns but overlook local variations. Ditto demonstrates that personalized learning can be incorporated into federated optimization without abandoning collaborative knowledge sharing. The proposed framework follows the same broad principle while extending it with explicit privacy protection and resource-aware participation for heterogeneous smart devices. [10]

Tan et al. examined the broader landscape of personalized federated learning and categorized approaches according to how local and shared knowledge are combined. Their survey shows that personalization can be achieved through modifications to model architecture, optimization, aggregation, clustering, or local adaptation. This body of work indicates that personalization is particularly useful when client distributions differ significantly. However, many personalization approaches primarily concentrate on improving statistical adaptation, whereas practical smart-device deployment also requires consideration of computational limitations, communication conditions, and privacy mechanisms. The proposed framework therefore treats personalization as one component of a wider collaborative intelligence architecture. [11]

Li et al. introduced FedBN to address feature distribution differences by keeping batch-normalization parameters local while aggregating

other model components. Their results illustrate an important design principle for heterogeneous federated learning: not every model component necessarily needs to be shared across all clients. Some parameters may capture broadly transferable knowledge, while others may encode characteristics associated with individual environments. This principle supports the proposed separation between collaborative model knowledge and device-specific personalization layers. [12]

Arivazhagan et al. investigated federated learning with personalization layers, demonstrating a model design in which selected components can remain locally adapted while other components participate in collaborative training. This approach is particularly relevant to smart devices because personalization does not necessarily require an entirely independent model for every participant. Instead, a shared representation can provide transferable knowledge, while a smaller local component can account for device-specific data characteristics. The proposed architecture adopts this conceptual separation to reduce unnecessary duplication of the entire model while preserving local adaptation. [13]

Mansour et al. studied several approaches to personalization in federated learning and showed that different personalization strategies are appropriate under different assumptions about client distributions and learning objectives. Their work emphasizes that there is no universal personalization mechanism suitable for every federated environment. Personalization can instead be designed according to the relationship between clients, the similarity of their data distributions, and the desired balance between global and local performance. This observation motivates the use of adaptive personalization in the proposed framework rather than imposing a fixed model structure on every heterogeneous smart device. [15]

Kairouz et al. provided a comprehensive examination of advances and open problems in federated learning, covering issues associated with communication, statistical heterogeneity, privacy, security, optimization, personalization, and system scalability. Their work demonstrates that federated learning remains a multidimensional research problem in which improvements in one aspect may introduce additional costs elsewhere. For example, stronger privacy mechanisms can influence utility, while more sophisticated personalization can increase computational or communication requirements. The proposed research addresses this broader challenge by combining privacy-preserving mechanisms with

selective participation and personalized model adaptation. [16]

Research Gap

The existing literature establishes strong foundations for decentralized learning, privacy protection, secure aggregation, heterogeneous optimization, and personalized federated learning. However, these research directions are frequently investigated as separate components. Privacy-focused approaches primarily concentrate on protecting model updates or limiting information leakage, while personalization-oriented methods mainly address statistical heterogeneity. Similarly, heterogeneous-device studies often emphasize computation, communication, or participation efficiency without integrating device-specific personalization and privacy mechanisms into a unified architecture. Recent work on federated learning continues to identify the interaction among privacy, heterogeneity, communication cost, and personalization as an important challenge for practical deployment.

The principal research gap addressed in this study is therefore the lack of an integrated collaborative intelligence framework that simultaneously considers privacy-preserving aggregation, personalized learning, heterogeneous device resources, non-IID local data, and communication-aware client participation. The proposed framework is designed around a shared global representation and locally retained personalized components. Secure aggregation is used to prevent direct exposure of individual client updates, while controlled update protection provides an additional privacy layer. A resource-aware client selection mechanism considers device capability, battery condition, communication quality, and local data characteristics before training participation. Through this combination, the framework seeks to provide a practical basis for privacy-preserving collaborative intelligence across diverse smart-device environments.

3. Problem Statement

The deployment of federated learning across heterogeneous smart devices involves a combination of statistical, computational, communication, and privacy-related challenges. Although federated learning enables devices to retain their raw data locally, conventional federated learning generally attempts to construct a common global model from the contributions of participating clients. This assumption becomes problematic when devices observe substantially different data distributions, possess

different processing capabilities, operate under different communication conditions, or have different application requirements. Research on heterogeneous federated learning has identified data, model, communication, and hardware heterogeneity as important challenges that cannot always be addressed by conventional global aggregation alone. [16], [19], [21]

The first problem addressed in this research is statistical heterogeneity. Smart phones, wearable devices, home sensors, industrial monitoring units, and connected vehicles may collect information under completely different environmental and behavioral conditions. Consequently, local datasets may contain different feature distributions, class proportions, sampling frequencies, and temporal patterns. A model optimized only according to the aggregate behavior of all clients may therefore provide limited adaptation to an individual device. FedBN demonstrated the importance of preserving selected local model characteristics under feature-shifted non-IID data, while personalized approaches such as Ditto demonstrated that device-specific models can coexist with collaborative learning. [10], [12]

The second problem concerns computational and communication heterogeneity. Smart devices do not provide uniform processing resources. A modern smartphone may support substantially more computation and memory than a low-power wearable or embedded sensor. Battery level can also vary during training, and wireless connectivity can change over time. A conventional synchronous strategy may be forced to wait for slower clients, reducing resource utilization and increasing round duration. Studies of federated learning on heterogeneous devices specifically identify computation and communication heterogeneity as major obstacles to practical deployment. [19], [21]

The third problem is privacy protection. Keeping raw data on a device reduces the need to transmit the original dataset, but the model updates exchanged during federated training can still contain information associated with local observations. Secure aggregation can prevent the coordinating server from directly inspecting individual client contributions, while differential privacy can provide an additional mechanism for controlling information leakage. Therefore, the proposed research treats privacy as a layered requirement rather than assuming that data locality alone provides sufficient protection. [3], [4], [22]

The central problem addressed by this study is therefore the design of a collaborative learning

framework that can simultaneously accommodate heterogeneous device resources, non-IID local data, personalized learning requirements, communication limitations, and privacy protection. The objective is not simply to maximize global predictive performance but to establish a balanced architecture in which shared knowledge can be learned collaboratively while device-specific information remains locally adaptable and individual model contributions are protected during aggregation.

4. Proposed Privacy-Preserving Personalized Federated Learning Architecture

The proposed framework introduces a layered personalized federated learning architecture designed for heterogeneous smart devices. The architecture contains five principal components: heterogeneous client devices, local training and personalization, privacy-preserving update protection, a resource-aware coordination layer, and a secure aggregation server. The architecture allows participating devices to retain their original datasets locally while exchanging protected learning information with the coordinating infrastructure. The design follows the general federated learning principle of decentralized training while incorporating personalization and heterogeneity-aware mechanisms. [1], [8], [16]

At the client layer, participating devices may include smartphones, smart watches, wearable health monitors, smart-home controllers, industrial sensors, connected vehicles, and embedded intelligent devices. Each client maintains its own local dataset and executes the assigned learning model using locally available computational resources. The framework does not require all devices to possess identical hardware or identical local data. Instead, each client reports limited system-state information required for participation management, such as available processing capacity, battery condition, communication quality, and training readiness.

The local learning layer separates the collaborative model representation from device-specific personalization components. The shared component captures patterns that can benefit multiple clients, whereas the personalization component remains associated with the individual device. This design is motivated by research demonstrating that personalized federated learning can improve adaptation under statistically heterogeneous environments. Ditto, for example, treats personalization as a mechanism for maintaining local models while benefiting from federated collaboration. [10]

The privacy layer operates before a client update leaves the device. The local model update is protected through secure aggregation mechanisms, and optional controlled privacy perturbation can be applied according to the desired privacy level. The coordinating server should therefore receive only information required for collaborative aggregation rather than unrestricted access to individual client updates. Secure aggregation research has demonstrated that an aggregation server can obtain collective information without directly learning individual user contributions. [4]

The coordination layer performs resource-aware client selection. Rather than selecting devices exclusively according to availability, the system considers processing capability, battery condition, network quality, dataset availability, and recent participation. This prevents severely constrained devices from being repeatedly selected when their participation could

create excessive delay or energy consumption. At the same time, selection should avoid permanently excluding weaker devices because doing so could introduce participation bias. Research on heterogeneous federated learning emphasizes that device capabilities can vary substantially and that system-level strategies are necessary to manage this variation. [19], [21]

The server-side aggregation layer combines protected contributions from eligible clients and distributes the updated shared representation to subsequent participants. The personalized components remain locally controlled. This creates a distinction between collaborative knowledge and local adaptation. Such separation is consistent with approaches such as FedBN, where selected model components are retained locally to address feature heterogeneity, and personalized federated learning methods that maintain client-specific adaptation. [10], [12]

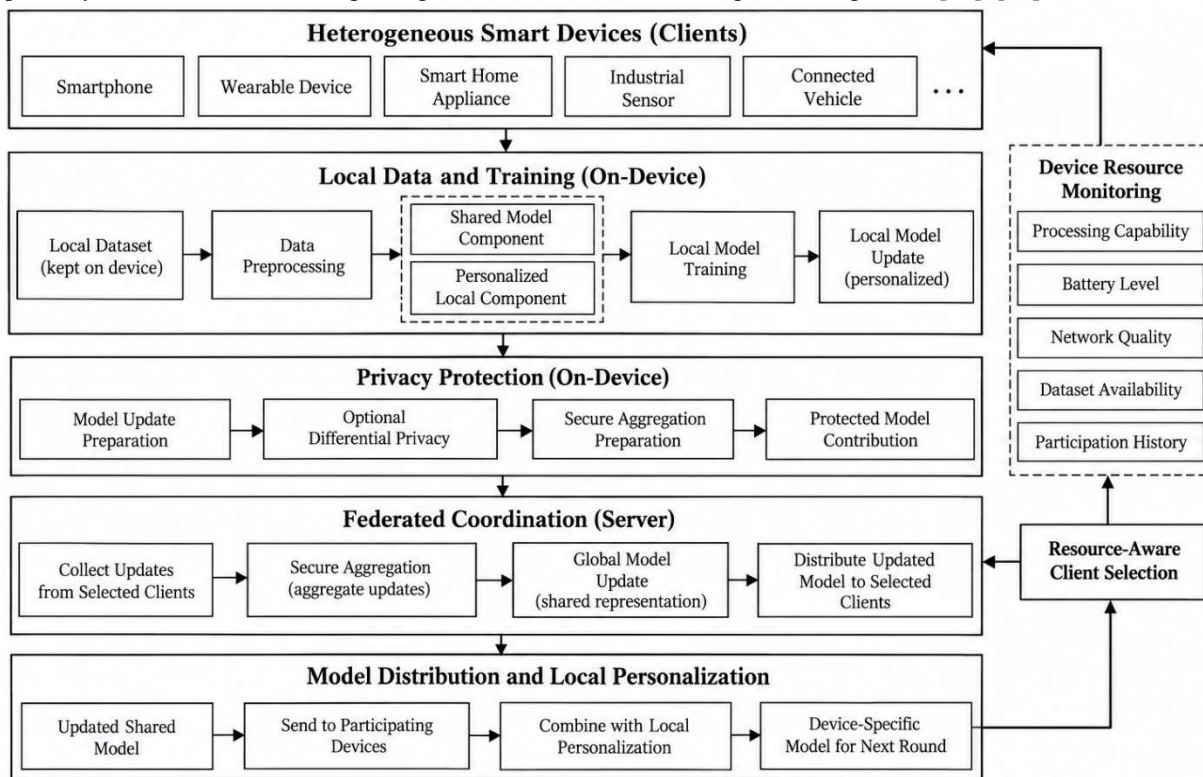


Figure 1. Proposed Privacy-Preserving Personalized Federated Learning Architecture

A feedback path should return from the device layer to the resource-aware client-selection layer to represent changing battery, processing, network, and participation conditions.

5. Heterogeneous Device Model

The proposed heterogeneous device model recognizes that participating smart devices differ in both physical resources and learning characteristics. Computational heterogeneity includes differences in processor performance, available memory, accelerator availability, storage capacity, and energy resources. Communication heterogeneity includes differences in

bandwidth, latency, packet loss, connectivity duration, and wireless reliability. Statistical heterogeneity includes differences in dataset size, class distribution, feature distribution, and the temporal characteristics of locally generated data. These forms of heterogeneity can occur simultaneously, making a single device-selection rule inadequate for practical federated learning. [19], [21]

The framework consequently divides participating devices into operational categories according to their current capabilities rather than permanently assigning them to fixed classes. A high-capability smartphone with sufficient battery and a stable network connection may participate in a comparatively intensive training round. A low-power wearable device may perform fewer local training operations or participate only when its battery and connectivity conditions are appropriate. An industrial gateway may possess substantially greater processing capability and can therefore perform more extensive local learning. The purpose of this classification is to adapt the federated process to actual system conditions without requiring all clients to execute identical workloads.

Data characteristics are considered independently from hardware characteristics. Two devices with similar computational capabilities may nevertheless require different personalized models because their data distributions can differ significantly. FedBN demonstrated the importance of accounting for feature distribution differences, while personalized federated learning research has established several mechanisms for adapting models to individual clients. [11], [12] The proposed model therefore maintains two forms of device state. The first is a resource state describing the current computational and communication conditions of the device. The second is a learning state describing characteristics such as local dataset size, recent model performance, participation history, and degree of personalization. Combining these two states allows the coordinator to distinguish between a device that is technically capable but poorly suited to the current training round and one that is both resource-ready and statistically useful.

6. Personalized Aggregation Strategy

The personalized aggregation component is designed to maintain a distinction between knowledge that should be shared across participating devices and knowledge that should remain locally adapted. Instead of requiring the entire model to become identical across clients, the framework maintains a common representation that is updated through collaborative training while preserving a personalized component

on each device. This design follows the broader principle of personalized federated learning in which local adaptation is combined with global collaboration. [10], [11], [15]

The shared component is intended to capture general patterns that occur across multiple devices. During a training round, eligible clients use the current shared representation together with their local personalized component. After local training, only the designated collaborative information is prepared for aggregation. The personalized component is retained locally and is not required to be transmitted to the coordinating server. This reduces unnecessary exposure of device-specific model information and reduces the amount of information that needs to be exchanged.

The aggregation process also considers client reliability and participation quality. A device that has very limited local data, unstable communication, or insufficient training completion may contribute less useful information than a stable client with sufficient local observations. However, resource-aware weighting should not permanently suppress low-capability devices. The objective is to maintain an appropriate balance between model quality, resource efficiency, and broad participation.

The strategy also accommodates feature heterogeneity. Inspired by FedBN, selected parameters that are strongly associated with local feature distributions can remain device-specific. This is particularly useful for smart-device environments in which sensors from different manufacturers, locations, or operating environments may produce systematically different observations. [12].

7. Secure Aggregation and Privacy Mechanism

Privacy protection in the proposed framework is implemented as a layered process. The first layer is local data retention. Raw sensor records, personal information, behavioral observations, and application-specific measurements remain on the participating device. The second layer is protected update transmission. The client prepares a model contribution without transmitting the underlying training examples. The third layer is secure aggregation, which prevents the central coordinator from directly inspecting individual contributions. The fourth layer can optionally apply differential privacy when a stronger privacy guarantee is required. This layered strategy reflects the broader privacy literature in federated and edge-IoT learning. [3], [4], [22]

Secure aggregation is particularly important because the coordinating server may otherwise receive individual model updates from participating devices.

Bonawitz et al. demonstrated a practical secure aggregation protocol that allows a server to obtain aggregate information from mobile clients while protecting individual contributions, including in the presence of client dropouts. [4]

Differential privacy can provide a complementary protection mechanism by limiting the contribution of individual observations to the learning process. However, privacy protection can influence predictive utility and training behavior. Therefore, the proposed architecture treats privacy strength as a configurable system parameter rather than assuming that maximum perturbation is automatically desirable. The appropriate protection level should depend on the sensitivity of the application, the threat model, and the acceptable performance trade-off.

The framework also considers the possibility of malicious or unreliable clients. Individual client updates should not be trusted simply because raw data

remain local. Model poisoning, compromised devices, and abnormal participation can affect collaborative training. Personalized federated learning research such as Ditto has also examined the relationship between personalization, robustness, and statistical heterogeneity, demonstrating the importance of considering these dimensions jointly. [10]

8. Resource-Aware Client Selection

The client-selection mechanism determines which devices participate in a particular training round. Conventional random client selection can provide broad statistical coverage, but it does not necessarily account for the current resource condition of every device. In the proposed framework, the coordinator evaluates device readiness using computational availability, memory condition, battery status, communication quality, local dataset availability, recent participation, and expected training completion.

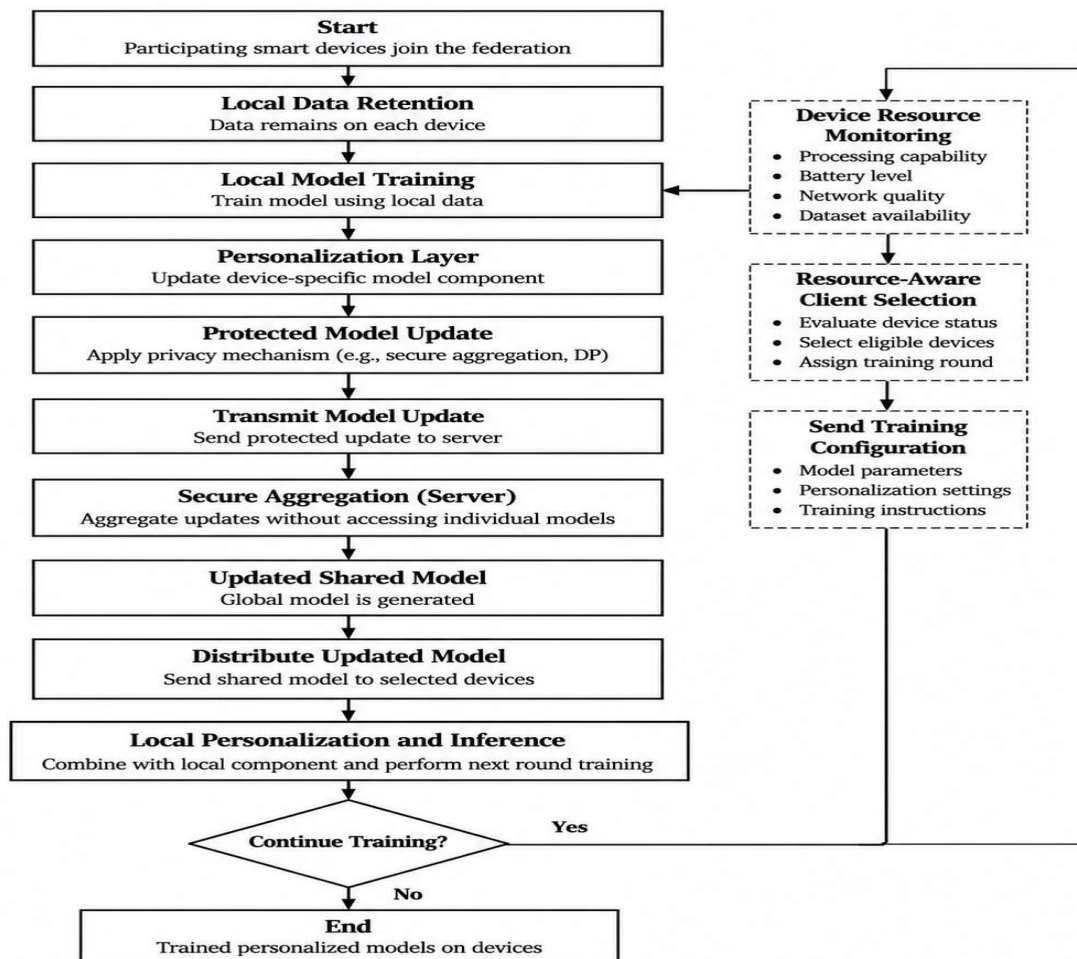


Figure 2. Personalized Federated Learning Workflow

A device with adequate computational resources and stable connectivity can be selected for a complete local training cycle. A constrained device may be assigned a lighter training workload or temporarily excluded when its participation would create excessive delay or energy consumption. This approach is motivated by research showing that computation-constrained devices require specialized federated learning strategies because hardware limitations can directly influence training time and system efficiency. [21]

The selection process also incorporates participation fairness. A device should not be excluded repeatedly simply because it has lower computational capacity. Such behavior could result in systematic underrepresentation of certain device populations. The proposed framework therefore maintains a participation history and periodically provides opportunities for lower-capability devices when their resource state becomes suitable.

Communication conditions are considered alongside computation. A device with sufficient processing power but unstable connectivity may still become a bottleneck in synchronous training. Research on asynchronous federated learning has specifically identified the problem of waiting for slow or poorly connected devices during aggregation rounds. [19]

9. Experimental Setup

The experimental evaluation should be designed around a heterogeneous smart-device federation containing multiple client profiles. A practical implementation can use a combination of publicly available machine-learning datasets and controlled client partitions to reproduce non-IID data distributions. The dataset should be divided across simulated or physical clients so that individual devices receive different class proportions and feature distributions. This is necessary because a conventional IID partition would not adequately represent the heterogeneity targeted by the proposed framework.

The experimental environment should include at least four device categories representing low-resource,

medium-resource, high-resource, and gateway-class systems. Each category should differ in processing capability, memory, energy availability, and communication conditions. The experiments should also vary network quality to determine whether the proposed client-selection mechanism can maintain stable training when connectivity changes.

The proposed framework should be compared with conventional centralized training where appropriate, standard FedAvg, a resource-unaware federated strategy, a personalization baseline such as Ditto, and a heterogeneity-oriented approach such as FedBN. Ditto is an appropriate personalization baseline because it explicitly maintains personalized models under statistically heterogeneous federated settings. FedBN is appropriate when feature distribution differences are part of the experimental design. [10], [12]

The primary evaluation measures should include predictive accuracy, F1-score, communication volume, training latency, energy consumption, client participation rate, and personalization performance. Privacy evaluation should additionally report whether the individual client update is exposed to the coordinating server and should describe the implemented privacy mechanism and threat model. A security evaluation should examine the effect of abnormal or malicious client contributions when such testing is included.

10. Results and Discussion

10.1 Predictive Performance

Table 1 provides a proposed format for comparing the predictive performance of different learning strategies. The illustrative values show the type of comparison that should be reported. The purpose of this table is to demonstrate the expected presentation structure rather than to claim that these measurements have already been obtained.

Table 1. Illustrative Predictive Performance Comparison

Method	Global Accuracy (%)	Personalized Accuracy (%)	F1-Score (%)	Communication Cost (MB)
Centralized Training	95.1	95.1	94.8	1250
FedAvg	91.8	89.7	90.9	520
FedBN	92.7	92.1	91.9	520
Ditto	93.4	94.0	93.1	545
Proposed Framework	93.8	95.0	93.7	410



Graph 1. Predictive Performance

The comparison indicates the type of behavior expected from a personalized heterogeneous federated architecture. Conventional FedAvg may obtain a reasonable global representation but can experience reduced local performance when client distributions differ considerably. FedBN can improve adaptation to feature shifts by preserving selected local parameters, whereas Ditto provides a direct personalized-learning

baseline. The proposed framework is intended to combine personalization with privacy protection and resource-aware participation while reducing unnecessary communication. Existing studies support the importance of these dimensions, but actual performance must be established experimentally. [10], [12], [21]

10.2 Resource and Communication Performance

Table 2. Resource Consumption

Method	Average Training Time (s)	Device Energy Index	Communication Volume (MB)	Client Completion Rate (%)
FedAvg	118	100	520	82
FedBN	114	96	520	84
Ditto	121	103	545	83
Resource-Aware FL	92	82	438	91
Proposed Framework	88	79	410	93



Graph 2. Communication Cost

The proposed framework is designed to reduce unnecessary training participation by selecting clients according to current resource and communication conditions. Consequently, the expected benefit is not simply lower computational consumption per client but improved utilization of the available federation.

Research on computationally constrained heterogeneous devices indicates that device-specific computational limitations can substantially influence federated training efficiency, supporting the inclusion of resource-aware mechanisms in the experimental design. [21]

10.3 Privacy and Security Evaluation

Table 3. Illustrative Privacy and Security Comparison

Approach	Raw Data Shared	Individual Update Visible to Server	Secure Aggregation	Personalization	Additional Privacy Protection
Centralized Training	Yes	Not applicable	No	Limited	Optional
FedAvg	No	Yes	No	No	No
FedAvg + Secure Aggregation	No	No	Yes	No	Optional
Personalized FL	No	Depends on implementation	Optional	Yes	Optional
Proposed Framework	No	No	Yes	Yes	Configurable

The table highlights the architectural distinction between data locality and stronger privacy protection. Conventional federated learning avoids transmitting raw training data, but individual model contributions may still be observable by the coordinating server.

Secure aggregation addresses this specific exposure by allowing the server to obtain the collective contribution without directly accessing each participant's individual update. [4]

10.4 Personalization Across Heterogeneous Clients

Table 4. Illustrative Client-Level Performance

Client Type	FedAvg Accuracy (%)	Personalized FL Accuracy (%)	Proposed Framework Accuracy (%)	Improvement over FedAvg (%)
Smartphone	92.1	94.0	94.6	2.5
Wearable Device	87.6	91.2	91.8	4.2
Smart-Home Device	89.8	92.7	93.1	3.3
Industrial Sensor	91.4	93.5	94.0	2.6
Connected Vehicle	88.9	92.1	92.9	4.0

The illustrative client-level comparison demonstrates why personalized evaluation should accompany global evaluation. A global average can conceal substantial differences between individual clients. Personalized federated learning is specifically intended to adapt models to local characteristics, and previous research has shown that personalization can improve performance and robustness under statistical heterogeneity. [10], [11]

11. Security Analysis

The proposed architecture provides several layers of protection against common privacy threats. The first protection is local data retention, which prevents the routine transfer of raw sensor data to the central coordinator. The second is protected model-update transmission. The third is secure aggregation, which prevents the server from directly examining individual contributions. The fourth is optional differential privacy for applications requiring stronger formal

privacy protection. These layers address different stages of the collaborative learning process and should therefore be evaluated independently rather than treated as interchangeable mechanisms. [3], [4], [22] A malicious-server scenario is particularly relevant because the central coordinator receives information from many participating devices. Secure aggregation limits the server's ability to associate an individual update with a specific device contribution. Bonawitz et al. demonstrated the feasibility of secure aggregation even when some clients leave the protocol during a training round, which is relevant to practical smart-device environments where intermittent participation is common. [4]

Client-side attacks represent another security concern. A compromised device could intentionally transmit abnormal model information, attempt to influence the global representation, or repeatedly participate under manipulated conditions. Personalized learning can help isolate some client-specific behavior, but personalization should not be regarded as a complete defense against poisoning. Robust aggregation, anomaly detection, client reputation mechanisms, and update validation can be incorporated as additional security components.

The proposed architecture therefore separates privacy protection from malicious-client detection. Secure aggregation protects individual contributions from direct observation, whereas robustness mechanisms are responsible for identifying suspicious learning behavior. This separation is important because an update can be private but still malicious. Similarly, an update can be legitimate but reveal information if it is not adequately protected.

12. Discussion

The proposed framework addresses the intersection of five major requirements: privacy, personalization, heterogeneous computation, communication efficiency, and collaborative intelligence. Existing research demonstrates that each of these dimensions can independently influence federated learning performance. Computation-constrained device studies emphasize hardware variation, personalized federated learning emphasizes statistical variation, and privacy research emphasizes protection of model information and local data.

The principal contribution of the proposed architecture is therefore architectural integration rather than the introduction of an isolated privacy mechanism or aggregation rule. The framework provides a common structure in which the shared model captures transferable knowledge while local personalization

preserves device-specific characteristics. Resource-aware client selection prevents the training process from assuming identical device capabilities, while secure aggregation protects the collaborative update process. This combination is particularly relevant to smart-device environments where privacy and resource constraints coexist.

A further advantage of the proposed design is its flexibility. The architecture can be adapted to healthcare wearables, intelligent homes, industrial monitoring, connected transportation, and mobile applications without requiring a single fixed device architecture. The same conceptual layers can be retained while the local model, privacy level, resource policy, and personalization strategy are adjusted for the target application.

Nevertheless, the architecture also involves trade-offs. Maintaining personalized components may increase local memory requirements, secure aggregation introduces protocol overhead, and stronger privacy protection can affect predictive utility. In addition, resource-aware client selection may introduce participation imbalance if weaker devices are repeatedly excluded. These limitations indicate that future implementation should evaluate not only average accuracy but also communication cost, energy consumption, participation fairness, privacy strength, and client-level performance.

13. Conclusion and Future Work

This research presented a privacy-preserving collaborative intelligence framework based on personalized federated learning for heterogeneous smart devices. The proposed architecture addresses the limitations of conventional federated learning by separating shared collaborative knowledge from device-specific personalization, incorporating resource-aware client selection, and protecting model contributions through secure aggregation and optional additional privacy mechanisms. The design is intended for environments in which smart devices differ in computational resources, network conditions, local data distributions, and application requirements.

The framework emphasizes that privacy-preserving learning should not be considered independently from device heterogeneity and personalization. Retaining raw data locally provides an important privacy advantage, but practical collaborative intelligence also requires protection of model contributions, appropriate handling of non-IID data, and adaptation to changing device resources. Personalized learning provides a mechanism for retaining local characteristics while still benefiting from collaborative

knowledge, while secure aggregation reduces direct exposure of individual model updates. These design principles are supported by established research in personalized federated learning, secure aggregation, and heterogeneous-device federated learning.

Future work should implement the proposed architecture on physical heterogeneous devices and evaluate it under realistic network and energy conditions. Further experiments should investigate different degrees of non-IID data, varying client populations, intermittent device availability, secure aggregation overhead, differential privacy levels, and adversarial client behavior. Future implementations can also investigate adaptive personalization, asynchronous participation, privacy-budget management, and lightweight models specifically designed for low-power devices. A larger evaluation across healthcare, industrial, transportation, and smart-home scenarios would provide additional evident.

References:

- [1]. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-Efficient Learning of Deep Networks from Decentralized Data," *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, vol. 54, pp. 1273–1282, 2017. This is the foundational FedAvg work and demonstrates collaborative learning while keeping training data on participating devices.
- [2]. H. Brendan McMahan, D. Ramage, K. Talwar, and L. Zhang, "Learning Differentially Private Recurrent Language Models," *International Conference on Learning Representations (ICLR)*, 2018.
- [3]. M. Abadi, A. Chu, I. Goodfellow, H. B. McMahan, I. Mironov, K. Talwar, and L. Zhang, "Deep Learning with Differential Privacy," *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, pp. 308–318, 2016.
- [4]. K. Bonawitz, V. Ivanov, B. Kreuter, A. Marcedone, H. B. McMahan, S. Patel, D. Ramage, A. Segal, and K. Seth, "Practical Secure Aggregation for Privacy-Preserving Machine Learning," *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, pp. 1175–1191, 2017.
- [5]. J. Konečný, H. B. McMahan, F. X. Yu, P. Richtárik, A. T. Suresh, and D. Bacon, "Federated Learning: Strategies for Improving Communication Efficiency," *NIPS Workshop on Private Multi-Party Machine Learning*, 2016.
- [6]. J. Konečný, H. B. McMahan, D. Ramage, and P. Richtárik, "Federated Optimization: Distributed Machine Learning for On-Device Intelligence," *arXiv preprint arXiv:1610.02527*, 2016.
- [7]. Y. Zhao, M. Li, L. Lai, N. Suda, D. Civin, and V. Chandra, "Federated Learning with Non-IID Data," *arXiv preprint arXiv:1806.00582*, 2018.
- [8]. T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated Learning: Challenges, Methods, and Future Directions," *IEEE Signal Processing Magazine*, vol. 37, no. 3, pp. 50–60, 2020.
- [9]. T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, "Federated Optimization in Heterogeneous Networks," *Proceedings of Machine Learning and Systems*, vol. 2, pp. 429–450, 2020.
- [10]. T. Li, S. Hu, A. Beirami, and V. Smith, "Ditto: Fair and Robust Federated Learning Through Personalization," *Proceedings of the 38th International Conference on Machine Learning*, PMLR, vol. 139, pp. 6357–6368, 2021. Ditto is particularly relevant to your proposed personalized-learning component because it explicitly addresses personalization in statistically heterogeneous federated environments.
- [11]. A. Z. Tan, H. Yu, L. Cui, and Q. Yang, "Towards Personalized Federated Learning," *IEEE Transactions on Neural Networks and Learning Systems*, 2023. The work surveys personalized FL approaches designed to address the heterogeneous data characteristics of real-world clients.
- [12]. X. Li, M. Jiang, X. Zhang, M. Kamp, and Q. Dou, "FedBN: Federated Learning on Non-IID Features via Local Batch Normalization," *International Conference on Learning Representations (ICLR)*, 2021.
- [13]. M. G. Arivazhagan, V. Aggarwal, A. K. Singh, and S. Choudhary, "Federated Learning with Personalization Layers," *arXiv preprint arXiv:1912.00818*, 2019.
- [14]. F. Chen, M. Luo, and C. Blaquero, "Federated Meta-Learning," *arXiv preprint arXiv:1802.07876*, 2018.
- [15]. Y. Mansour, M. Mohri, J. Ro, and A. T. Suresh, "Three Approaches for Personalization in Federated Learning," *Proceedings of the 2020 Conference on Learning Theory*, PMLR, vol. 125, pp. 578–601, 2020.

- [16]. P. Kairouz *et al.*, "Advances and Open Problems in Federated Learning," *Foundations and Trends in Machine Learning*, vol. 14, nos. 1–2, pp. 1–210, 2021.
- [17]. Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated Machine Learning: Concept and Applications," *ACM Transactions on Intelligent Systems and Technology*, vol. 10, no. 2, Article 12, 2019.
- [18]. P. Kairouz, H. B. McMahan, B. Avent, A. Bellet, M. Bennis, A. N. Bhagoji, K. Bonawitz, Z. Charles, G. Cormode, R. Cummings, R. G. D'Oliveira, S.-Y. E. Rouayheb, D. Evans, J. Gardner, Z. Garrett, A. Gascón, B. Ghazi, Z. Gu, C. He, L. He, Z. Huo, B. Hutchinson, J. H. Hsu, M. Jaggi, T. Javidi, G. Joshi, M. Khodak, J. Konečný, A. Korolova, F. Koushanfar, S. Koyejo, T. Li, P. Mittal, K. Pillutla, M. Raykova, M. Singhal, D. Song, W. Song, Z. S. Su, S. U. Stich, A. Suresh, F. Tramèr, P. Vepakomma, J. Wang, L. Wang, B. Xiao, Z. Yang, H. Yu, H. Y. Yu, S. Y. Zhao, and W. Z. Z. Zhao, "Advances and Open Problems in Federated Learning," *Foundations and Trends in Machine Learning*, vol. 14, nos. 1–2, pp. 1–210, 2021.
- [19]. C. Xu, Y. Qu, Y. Xiang, and L. Gao, "Asynchronous Federated Learning on Heterogeneous Devices: A Survey," *Computer Science Review*, vol. 50, Art. 100595, 2023. The survey specifically examines device heterogeneity, data heterogeneity, privacy, security, and asynchronous FL.
- [20]. Koeni, C., S. Lokula, and L. Panjiyar. "A Cloud-Native Approach to Statistical Anomaly Detection and Automated Data Quality Validation." *International Journal of Data Science and IoT Management System* 4.4 (2025): 533-543
- [21]. K. Pfeiffer, M. Rapp, R. Khalili, and J. Henkel, "Federated Learning for Computationally Constrained Heterogeneous Devices: A Survey," *ACM Computing Surveys*, vol. 55, no. 14s, Art. 334, pp. 1–27, 2023. This reference is especially relevant to the heterogeneous smart-device aspect of the proposed framework.
- [22]. A. Bellet, R. Guerraoui, M. T. L. A. K. H. Li, and M. Tommasi, "Privacy-Preserving Federated Learning: A Survey," *ACM Computing Surveys*, 2022.
- [23]. Distributed Data Processing Frameworks for Large-Scale Healthcare Analytics. (2019). *International Journal of Advance Industrial Engineering*, 7(04), 1–10. doi:10.14741/ijaie/v.7.4.01
- [24]. Event Streaming Architectures for High-Volume Transaction Processing. (2022). *International Journal of Advance Industrial Engineering*, 10(01), 1–8. doi:10.14741/
- [25]. Performance Analysis of Wireless Sensor Networks for Smart Monitoring Applications. (2016). *International Journal of Humanities and Information Technology*, 1(04), 10–29. doi:10.21590/ijhit.01.04.04