Chapter 7

Auction-Driven Utility Maximization for Multiple FL Systems in WBAN-based Smart Healthcare

7.1 Introduction

With the advanced sensing capabilities of wearable health sensors, WBANs are increasingly used in smart healthcare for real-time monitoring applications, such as self-management interventions, digital diabetes coaching, and obesity prevention, thereby enhancing users' quality of life [13, 77, 149]. Additionally, this health data can aid in developing personalized ML models for various healthcare applications [2]. However, traditional ML methods require WBAN users to send their data to a central server for model training, which raises privacy concerns and regulatory challenges, leading to data islands. To this, FL paradigm offers a solution by enabling multiple WBAN users to collaboratively train a model while maintaining privacy, with the help of a server [150]. In this process, eligible WBAN users first download a global model from a server, per-

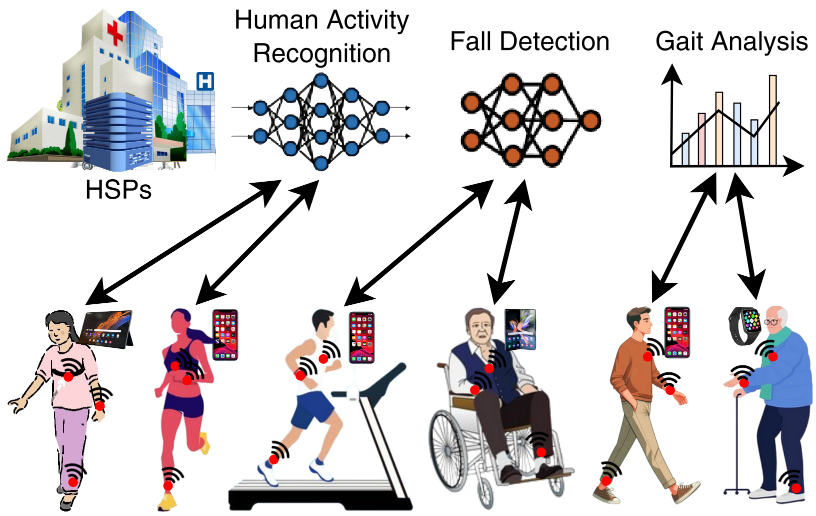


Fig. 7.1. An instance of multiple FL models in smart healthcare.

form local training with their health data, and upload the trained model parameters back to the server, which aggregates them into a new global model [25]. Although only model parameters are transmitted to the aggregation server, the risk of privacy breaches remains, as adversaries might infer original data from these parameters. To mitigate this privacy issue, DP strategies have been employed in the FL framework [29, 72].

Typically, a model owner, such as a HSP, aims to train an ML model using WBAN users' health data. However, WBAN users are often reluctant to participate in FL process due to privacy concerns and the costs of using their own computational and communication resources for FL training [21]. *Therefore, HSPs must offer incentives to WBAN users to cover these costs and encourage participation.* Current research, however, primarily focuses on single FL model, addressing aspects such as incentive mechanisms, user selection based on amount of data [29, 70], privacy budget [71–73], training latency [79], training cost [74], model accuracy [75], and reputation [26, 76]. Despite this, single FL model engage only a subset of WBAN users, minimizing the impact of stragglers but leaving other users idle, which reduces overall efficiency [22]. In contrast, multiple FL can enhance efficiency by simultaneously training several models using diverse, decentralized health data from WBAN users [22]. As demand for MLpowered smart healthcare applications grows and diversifies, training multiple FL models becomes more feasible. For instance, multiple ML models, such as *human activity recognition*, *fall detection*, and *gait analysis* for monitoring the mobility of Parkinson's patients [23, 24], can be trained concurrently using the FL paradigm to address privacy concerns, as illustrated in Fig. 7.1. Similarly, WBAN data on hygiene and COVID-19 can be utilized through FL for health services such as *personalized recommendations*, *trending topic prediction*, and *COVID-19 detection* [80].

Generally, different FL models have varying learning requirements (e.g., model architecture, training deadlines) and economic values, where the quality of local health data (e.g., amount of health data, model accuracy, and reputation) from WBAN users is crucial for FL performance [25]. Additionally, WBAN users are heterogeneous, with varying computation and communication capabilities, privacy needs, and data quality, which leads to differing learning costs. Furthermore, unreliable WBAN users may, intentionally or unintentionally, submit low-quality models that can negatively impact the global FL model [26]. Thus, selecting reliable WBAN users based on their attributes and FL model requirements is essential for the successful completion of FL training and for maximizing economic value (discussed in Section 7.2). However, the challenges of incentivizing WBAN users and selecting reliable users are interdependent and cannot be addressed separately [26]. To overcome this, auction-based incentive mechanisms for user selection have been proposed that guarantee computational efficiency, truthfulness, and individual rationality [70, 75]. In contrast to contract theory and Stackelberg game-based mechanisms [26, 76–78], auction mechanisms allow WBAN users to report their true costs and provides a natural way to balance supply and demand by setting appropriate prices and selecting suitable users [21]. However, existing auction-based approaches are not directly applicable to our proposed work, as they do not consider multiple FL models [70, 75]. Additionally, research on incentive mechanisms for multiple FL models [21, 25, 27, 28] primarily focuses on the amount of data and the training costs, while neglecting critical factors such as privacy requirements and user reputations in smart healthcare. Moreover, these works fail to consider the criticality of health data, but it is important to address this factor as it significantly affects WBAN users' willingness to share their data [29].

Motivated by the aforementioned scenario, this chapter presents an incentive framework designed to select reliable WBAN users with high-quality health data for differentially private multiple FL models in smart healthcare. Moreover, an incentive mechanism is devised, and an optimization problem is formulated to maximize system utility, incorporating a cost model that includes data collection, computation, communication, and privacy, as an NP-hard problem. An efficient auction-based algorithm is proposed to address the formulated problem that integrates factors such as local model accuracy, user reputation, and data volume. Additionally, a reputation scheme based on subjective logic models is employed as a key metric to evaluate user reliability [26,151]. Furthermore, the effectiveness of the proposed approach is evaluated through real-world data analysis across various settings. *To the best of our knowledge, this work is among the first to consider incentive mechanism for selecting reliable WBAN users for multiple FL models with DP in smart healthcare systems.* In summary, the contributions of this chapter are as follows:

- Propose a multiple FL model framework for smart healthcare applications that leverages health data from diverse WBAN users to train multiple ML models while addressing their privacy requirements.
- Formulate a utility minimization problem, incorporating a cost model that includes data collection, computation, communication, and privacy, altogether as an NP-hard.
- Propose an efficient **A**uction-**D**riven incentive mechanism for multiple **FL** (ADFL) models, designed for selecting WBAN users across multiple FL models, integrating local accuracy, reputation, and amount of health data. Furthermore, offers

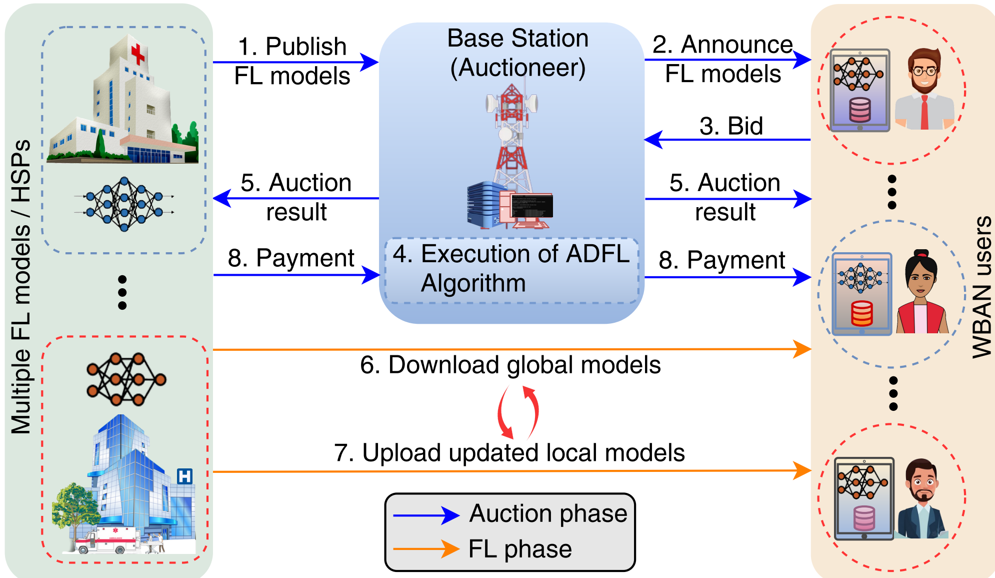


Fig. 7.2. Incentivised WBAN users selection for multiple FL.

theoretical analyses of auction properties, including individual rationality, incentive compatibility, budget feasibility, and computational efficiency.

- Find the efficacy of the proposed algorithm by achieving an average improvement of 15.9% and 18.08% in utility compared to state-of-the-art works through extensive simulations and analysis on real-world health data.

7.2 System Model and Problem Formulation

We consider a smart healthcare system in which J FL models, denoted as $\mathbb{J} = \{1, \ldots, j, \ldots, J\}$, each owned by different HSPs, are simultaneously trained using a set of P WBAN users, with the assistance of a BS. The BS facilitates communication between WBAN users and HSPs and conducts auction (i.e., execution of the ADFL algorithm) to select WBAN users for various FL models and determine their respective payments. To initiate the training of each FL model, HSPs first announce their models via the BS (points 1-2), as shown in Fig. 7.2. Afterward, WBAN users submit their bidding details (point 3), including computing resources, health data availability, privacy needs, and costs (discussed in the following subsections). The BS then conducts the auction to determine the selected users and their corresponding payments (point 4). Consequently, the BS announces the auction results to both WBAN users and HSPs (point 5). The selected WBAN users then perform FL model training[1], which involves three steps: (i) each HSP initializes and broadcasts the global model parameters to their respective WBAN users; (ii) WBAN users perform local model training using their own health data; and (iii) HSPs aggregate[2] the local model parameters from their respective WBAN users to obtain a global model. Finally, HSPs provide payments to the respective WBAN users for participating in the FL process (point 8). Despite only model parameters being transmitted to the aggregation server, the risk of privacy breaches remains, as adversaries might potentially infer original health data from these parameters. To address this concern, we adopt FL with DP to ensure that WBAN users' sensitive health data remains private, as discussed in the following.

7.2.1 Multiple FL with DP Model

We assume that each WBAN user possesses various types and quantities of health datasets, allowing them to participate in the training of multiple FL models independently. Let $\mathbb{P} = \{1, \ldots, p, \ldots, P\}$ be the set of WBAN users, each with health datasets for training FL models, denoted as $\{\mathscr{A}_{p,j}^{\mathrm{i}}\}_{\mathrm{i}=1}^{\mathrm{I}_{p,j}}$, where $\mathrm{I}_{p,j}$ represents number of health data samples and $\mathscr{A}_{p,j}^{\mathrm{i}}$ denotes the i^{th} health data sample. Based on FL model requirements, WBAN users perform iterative local model training with the objective of minimizing task-specific loss functions, such as cross-entropy, mean squared error, and

[1]This chapter primarily focuses on incentive design and reliable WBAN user selection for multiple FL models with DP in smart healthcare, rather than on the actual training of FL models (points 6-7), which we plan to explore in our future work.

[2]Each HSP operates a computing server that functions as the global model aggregator for its respective FL model [81].

log-likelihood [145]. Local training of FL model j by WBAN user p is defined as follows:

$$\min_{\mathbf{S}_{p,j}} \mathrm{F}_{p,j}(\mathbf{S}_{p,j}) := \min_{\mathbf{S}_{p,j}} \frac{1}{\mathrm{I}_{p,j}} \sum_{\mathrm{i}=1}^{\mathrm{I}_{p,j}} \mathrm{Q}(\mathbf{S}_{p,j}, \mathscr{A}^{\mathrm{i}}_{p,j}), \tag{7.1}$$

where $\mathbf{S}_{p,j}$ represents the local model parameters. Following the training, each WBAN user transmits its model parameters to the HSP for aggregation. Although FL enables local data processing and transmits only model parameters to the server for aggregation, still privacy vulnerabilities persist through shared model parameters. In particular, HSPs may extract WBAN users' training health data from the transmitted parameters or gradients [71]. Moreover, shared model parameters may be eavesdropped during transmission, potentially leading to reconstruction of WBAN users' original health data.

To overcome the above-mentioned limitations, we propose a framework for training multiple FL models that employs a DP-based approach to ensure privacy protection without significantly compromising local model accuracy. Specifically, each WBAN user employs the Gaussian DP mechanism to ensure privacy by adding Gaussian noise to the local loss function $\mathrm{F}_{p,j}(\mathbf{S}_{p,j})$ as follows: $\mathcal{Y} := \mathrm{F}_{p,j}(\mathbf{S}_{p,j}) + \mathscr{J}(0, \mathscr{e}^2_{p,j}\mathbf{I}_{\mathrm{d}})$, where $\mathscr{e}_{p,j}$ is the noise scale[3], and d is the dimension of the local model [29, 71]. The Gaussian mechanism $\mathcal{Y}$ achieves $(\epsilon_{p,j}, \mathrm{f}_{p,j})$-DP if $\mathrm{f}_{p,j} \in (0, 1)$ with $\epsilon_{p,j} > 0$, where $\epsilon_{p,j}$ is the privacy budget, and $\mathrm{f}_{p,j}$ is the bound for which the privacy guarantee may not hold [71]. If $\epsilon_{p,j}$ is too small, more noise needs to be added, preventing the model training from converging and achieving lower accuracy. Conversely, if $\epsilon_{p,j}$ is too large, less noise is required to achieve the privacy budget, resulting in higher accuracy; however, potentially compromising WBAN user's privacy. Therefore, privacy budget should adhere to the following constraint: $\epsilon_j^{min} \leq \epsilon_{p,j} \leq \epsilon_j^{max}$. Consequently, we define the relationship between local model accuracy $\mathrm{a}_{p,j}$ and privacy budget $\epsilon_{p,j}$. For generality, we directly adopt a continuous and reversible concave function to model this relationship, which is

[3] A detailed explanation of how to determine the appropriate noise scale to guarantee $(\epsilon_{p,j}, \mathrm{f}_{p,j})$-DP is provided in [71].

expressed as follows [73]:

$$\mathrm{a}_{p,j} = \top \epsilon_{p,j}, \tag{7.2}$$

where $\top$ is the conversion parameter from privacy budget to accuracy. Additionally, security techniques such as homomorphic encryption mechanisms can be applied to further prevent privacy leakage[4] during model parameter transmission [29].

To design an effective incentive mechanism for FL, it is crucial to consider the costs associated with participating in the FL process, as discussed below.

7.2.2 Cost Model

The cost model of WBAN users consists of data collection, privacy, computation, and communication costs, as discussed in the following.

7.2.2.1 Data collection cost

WBAN users must first collect health data samples to participate in FL model training, which incurs costs due to energy consumption during data collection. These costs arise not only from energy consumption but also from deploying WBANs, including body sensors and local devices (e.g., mobile phones, laptops). Let $\zeta_{p,j} > 0$ be the unit cost incurred by WBAN user p to collect health data for training FL model j. Then, the cost for WBAN user p to collect $\mathrm{I}_{p,j}$ health data samples is defined as follows [70]:

$$\mathscr{C}_{p,j}^{data} = \zeta_{p,j} \mathrm{I}_{p,j}. \tag{7.3}$$

7.2.2.2 Privacy cost

The privacy cost refers to the compensation sought by a WBAN user for potential economic losses resulting from privacy leaks. Additionally, a larger privacy budget

[4] Although integrating security technique such as homomorphic encryption with DP is outside the scope of this chapter, it can be achieved using the methods described in our previous work [29].

$\epsilon_{p,j}$ indicates a lower level of privacy protection, which leads to a higher privacy cost. Specifically, we assume that the privacy cost for WBAN user p is proportional to its privacy budget $\epsilon_{p,j}$, reflecting the privacy protection level, as done in [73]. Therefore, privacy cost for WBAN user p in training FL model j is defined as follows:

$$\mathscr{C}_{p,j}^{pri} = \xi_{p,j}\epsilon_{p,j}, \tag{7.4}$$

where $\xi_{p,j}$ is the unit privacy cost, which represents how much the WBAN user cares about its privacy loss.

7.2.2.3 Computation cost

Let $\mathrm{g}_{p,j}$ be the computing resources (CPU frequency) of WBAN user p for training FL model j, and let $\mathscr{w}_j$ be the number of CPU cycles required to compute one health data sample. The computation time for a local iteration is given by $\mathrm{M}_{p,j}^{cmp} = \mathscr{w}_j \mathrm{I}_{p,j}/\mathrm{g}_{p,j}$, while the energy consumption for WBAN user p is defined as $\mathrm{X}_{p,j}^{cmp} = \varepsilon_p \mathscr{w}_j \mathrm{I}_{p,j}\mathrm{g}_{p,j}^2$ [26], where ε_p represents the effective capacitance of WBAN user p's computing chip. Consequently, the cost incurred by WBAN p for local model computation is defined as follows [4, 29]:

$$\mathscr{C}_{p,j}^{cmp} = \pi_{p,j}\varepsilon_p^2 \mathscr{w}_j^2 \mathrm{I}_{p,j}^2 \mathrm{g}_{p,j}^4, \tag{7.5}$$

where $\pi_{p,j}$ represents the unit cost of computation energy. The value of $\pi_{p,j}$ directly depends on the criticality [4], denoted as $\mathscr{P}_{p,j}$, of WBAN user p, defined as follows:

$$\mathscr{P}_{p,j} = \sum_{i=1}^{\mathrm{I}_{p,j}} \mathscr{m}_{p,j}^{\mathrm{i}} \mathscr{I}_{p,j}^{\mathrm{i}}, \tag{7.6}$$

where $\mathscr{m}^{\mathrm{i}}_{p,j}$ denotes the medical criticality and $\mathscr{I}^{\mathrm{i}}_{p,j}$ represents the health severity index of health data $\mathscr{A}^{\mathrm{i}}_{p,j}$. Mathematically, $\mathscr{I}^{\mathrm{i}}_{p,j}$ is defined as follows [3, 4]:

$$\mathscr{I}^{\mathrm{i}}_{p,j} = \left| \frac{(\mathscr{Q}^{\mathrm{i}}_{up} - \mathscr{A}^{\mathrm{i}}_{p,j})^2 - (\mathscr{A}^{\mathrm{i}}_{p,j} - \mathscr{Q}^{\mathrm{i}}_{low})^2}{(|\mathscr{Q}^{\mathrm{i}}_{up}| + |\mathscr{Q}^{\mathrm{i}}_{low}|)^2} \right|, \tag{7.7}$$

where $\mathscr{Q}^{\mathrm{i}}_{low}$ and $\mathscr{Q}^{\mathrm{i}}_{up}$ are lower and upper bounds of health data under normal conditions.

7.2.2.4 Transmission cost

BS, which operates on Orthogonal Frequency-Division Multiple Access (OFDMA) technology, facilitates the transmission of local model parameters between WBAN users and HSPs[5]. The data transmission rate between WBAN user p and HSP for FL model j is denoted as $\mathscr{r}_{p,j}$, and it is affected by factors such as channel conditions, packet retransmission, transmission power, and outage probability [152]. Let $\mathscr{U}_j$ be the size of the local model, which remains constant for all WBAN users participating in the training of FL model j. Thus, the transmission time for local model of size $\mathscr{U}_j$ is defined as $\mathrm{M}^{trs}_{p,j} = \mathscr{U}_j / \mathscr{r}_{p,j}$. Therefore, the total time taken by WBAN user p in one global iteration of FL model j is defined as follows:

$$\mathcal{T}_{p,j} = \varsigma_{p,j} \log\left(1/\mathrm{a}_{p,j}\right) \mathrm{M}^{cmp}_{p,j} + \mathrm{M}^{trs}_{p,j}, \tag{7.8}$$

where $\varsigma_{p,j} \log\left(1/\mathrm{a}_{p,j}\right)$ represents the number of local iteration required to achieve a local accuracy of $\mathrm{a}_{p,j}$, and $\varsigma_{p,j}$ is a positive parameter that depends on $\mathrm{F}_{p,j}(\mathbf{S}_{p,j})$ [153]. Additionally, WBAN user p should complete a global iteration within a specified deadline, i.e., $\mathcal{T}_{p,j} \leq \mathcal{T}^{max}_j$. According to [26], the energy consumption for WBAN user p, with transmission power ϖ_p, to transmit the local model in a global iteration of FL model j is given by $\mathrm{X}^{trs}_{p,j} = \varpi_p \mathrm{M}^{trs}_{p,j}$. Therefore, the transmission cost for WBAN user p

[5]Bandwidth allocation is beyond the scope of this chapter; nevertheless, this can be done using the methods provided in [15].

participating in the training of FL model j is defined as follows:

$$\mathscr{C}_{p,j}^{trs} = \mathfrak{w}_p \mathrm{X}_{p,j}^{trs}, \tag{7.9}$$

where $\mathfrak{w}_p$ is the unit cost of transmission energy.

Therefore, the total cost for WBAN user p to participate in the training of FL model j is expressed as follows:

$$\mathbb{C}_{p,j} = \mathscr{C}_{p,j}^{data} + \mathscr{C}_{p,j}^{pri} + \varsigma_{p,j} \log\left(\frac{1}{\mathrm{a}_{p,j}}\right) \mathscr{C}_{p,j}^{cmp} + \mathscr{C}_{p,j}^{trs}. \tag{7.10}$$

The cost model primarily focuses on the expenses incurred by WBAN users, without considering their trustworthiness or reliability—a crucial factor for effective FL implementation, as discussed further below.

7.2.3 Reputation Model

The trustworthiness of a WBAN user is crucial, as it reflects their reliability in completing FL model training, which directly affects the overall performance of the global model. Trustworthy users who provide reliable and high-quality training data contribute to higher local model accuracy, thereby enhancing overall FL performance [26]. Inspired by [26, 76], we employ a subjective logic-based reputation scheme to evaluate the trustworthiness or reliability of WBAN users. This scheme consists of two components: the fact space $\mathscr{L}_{p,j}$ and the conceptual space $Ł_{p,j}$. The fact space is defined as $\mathscr{L}_{p,j} = \{\mathfrak{a}_{p,j}, \mathrm{t}_{p,j}\}$, where $\mathfrak{a}_{p,j}$ and $\mathrm{t}_{p,j}$ represent number of positive and negative interactions between the WBAN user and the HSP, respectively. If a WBAN user fails to complete an FL training iteration within the deadline, or if the local models they transmit are not useful[6], it is considered a negative interaction, and vice versa [26]. Meanwhile, the conceptual space is expressed as $Ł_{p,j} = \{z_{p,j}, d_{p,j}, \mathfrak{u}_{p,j}\}$, where $z_{p,j}, d_{p,j},$ and $\mathfrak{u}_{p,j}$

[6]Determining the usefulness of a local model is beyond the scope of this chapter; however, this can be achieved using the methods provided in [26, 76].

denote belief, disbelief, and uncertainty, respectively, and are defined as follows [76]:

$$\begin{cases} z_{p,j} = (1 - \mathfrak{u}_{p,j})(\frac{\mathfrak{a}_{p,j}}{\mathfrak{a}_{p,j}+\mathfrak{t}_{p,j}}), \\ d_{p,j} = (1 - \mathfrak{u}_{p,j})(\frac{\mathfrak{t}_{p,j}}{\mathfrak{a}_{p,j}+\mathfrak{t}_{p,j}}), \\ z_{p,j}, d_{p,j}, \mathfrak{u}_{p,j} \in [0,1], \\ z_{p,j} + d_{p,j} + \mathfrak{u}_{p,j} = 1. \end{cases} \tag{7.11}$$

The reputation of the HSP that owns FL model j for WBAN user p in the current iteration is defined as follows:

$$\mho'_{p,j} = \text{ß}_1 z_{p,j} - \text{ß}_2 d_{p,j} - \text{ß}_3 \mathfrak{u}_{p,j}, \tag{7.12}$$

where $\text{ß}_1, \text{ß}_2$, and ß_3 are the predetermined weights [76].

A WBAN user's trustworthiness or reliability evolves over time, and their reliability in interactions with HSPs changes. To account for historical reputation, we introduce a freshness parameter f in the subjective logic model. Following [76], the current iteration's reputation is assigned a higher weight compared to the historical reputation, capturing the impact of recent performance on overall reputation. By considering both the current reputation $\mho'_{p,j}$ and the historical reputation $\mho''_{p,j}$, the final reputation of the HSP that owns FL model j for WBAN user p is redefined as follows:

$$\mho_{p,j} = f \times \mho'_{p,j} + (1 - f) \times \mho''_{p,j}. \tag{7.13}$$

In the next subsection, we describe the auction-based incentive mechanism for selecting WBAN users.

7.2.4 Auction Model

This section presents the procedure of the proposed auction, the utility calculation for both WBAN users and HSPs, and the constraints associated with the auction mechanism.

7.2.4.1 Overall auction procedure

In the proposed auction, each HSP initially announces its FL model requirements, including FL training deadline ($\mathcal{T}_j^{max}$) and privacy budget range ($\epsilon_j^{min}, \epsilon_j^{max}$), to all WBAN users via the BS. Each WBAN user then submits their bid, which includes details such as available computing resources $\mathrm{g}_{p,j}$, number of health data samples $\mathrm{I}_{p,j}$, privacy budget $\epsilon_{p,j}$, and reported cost $\hat{\mathbb{C}}_{p,j}$. Once the BS receives all WBAN users' bid, it initiates the auction process to select WBAN users for each FL model training[7]. Moreover, we define a binary decision variable $\mathrm{m}_{p,j}$ to represent the selection of WBAN users for training the FL model, where $\mathrm{m}_{p,j} = 1$ indicates that WBAN user p is chosen for FL model j, otherwise $\mathrm{m}_{p,j} = 0$. Subsequently, the BS determines the payments for the selected WBAN users.

7.2.4.2 Utility calculation

A WBAN user's utility depends on whether it is selected to participate in the training of an FL model. If a WBAN user is chosen, its utility is defined as the difference between the payment $pay_{p,j}$ (determined by the proposed ADFL algorithm) received for participating in the training of the FL model and the associated costs; otherwise, the utility is 0, i.e.,

$$\Re_{p,j} = \begin{cases} pay_{p,j} - \mathbb{C}_{p,j}, & \text{if } \mathrm{m}_{p,j} = 1; \\ 0, & \text{otherwise.} \end{cases} \tag{7.14}$$

WBAN users with high reputations and abundant health data samples contribute high-quality local models, thereby improving the overall quality of the global model. Therefore, inspired by [77], we define the valuation of WBAN user p's local model based on factors such as accuracy, reputation, and the quantity of health data samples used

[7] Each WBAN user can participate in at most one FL model training; however, multiple WBAN users can participate in a single FL model training.

in model training, reflecting the HSP's satisfaction[8], as follows:

$$\mathbb{W}_{p,j} = \mathscr{y}\Big(\tan^{-1}\Big((\mathscr{w}_1 \mathrm{a}_{p,j} + \mathscr{w}_2 \mho_{p,j})\, \mathrm{I}_{p,j}\Big)\Big), \tag{7.15}$$

where $\mathscr{y} > 0$ is the conversion parameter for model performance into profit, $\mathscr{w}_1 > 0$ is the conversion coefficient for accuracy, and $\mathscr{w}_2 > 0$ is the conversion coefficient for reputation[9]. It is clear from Eq. (7.15) that higher accuracy and high reputation with abundant health data increase the satisfaction of HSP. Thus, the utility of HSP for its FL model j is calculated as the difference between the valuation and the payment to WBAN user p, representing the profit obtained when WBAN user p joins the training of its FL model. Mathematically, this can be expressed as follows:

$$\mathcal{U}_{p,j} = \mathbb{W}_{p,j} - pay_{p,j}. \tag{7.16}$$

WBAN users generally exhibit selfish behavior, prioritizing their own utility. As a result, they may manipulate their true costs to maximize their utility. Therefore, it is crucial to ensure that the auction mechanism encourages truthful bidding by WBAN users and satisfies the following properties:

Individual rationality: All the selected WBAN users can obtain non-negative utilities, i.e.,

$$\Re_{p,j} \geq 0. \tag{7.17}$$

Incentive compatibility (truthfulness): Each WBAN user with true cost $\mathbb{C}_{p,j}$ can

---

8 We adopt the tangential function in Eq. (7.15) because it effectively maps users' accuracy and reputation to their overall contribution in the FL system [10, 77]. Its gradual growth captures the diminishing impact of increasing accuracy and reputation, making it more suitable than exponential or additive-based functions, which fail to account for this diminishing effect.

9 The conversion coefficient for accuracy and reputation must be greater than 0, with no strict upper limit [77]. However, due to the use of the inverse tangential function, increasing the conversion coefficient beyond a certain point leads to diminishing returns, as the function asymptotically stabilizes.

only maximize its utility by truthfully reporting its true cost, i.e.,

$$\Re_{p,j}(\mathbb{C}_{p,j}) \geq \Re_{p,j}(\hat{\mathbb{C}}_{p,j}), \forall\ \hat{\mathbb{C}}_{p,j} \neq \mathbb{C}_{p,j}. \tag{7.18}$$

Since WBAN users' payments depend on their reported costs, they may misreport their costs if it results in higher payments. However, incentive compatibility (truthfulness) ensures that, under the auction mechanism, they are motivated to report their costs truthfully. This means WBAN users achieve the highest utility by truthfully reporting their costs [75].

Budget feasibility: Total payment to selected WBAN users must not exceed the budget allocated for FL model j, i.e.,

$$\sum_{p \in \mathbb{P}} \mathrm{m}_{p,j} pay_{p,j} \leq \mathcal{B}_j, \tag{7.19}$$

where $\mathcal{B}_j$ represents the budget for FL model j.

7.2.5 Problem Formulation

In this chapter, HSPs strive to maximize their utility by selecting reliable WBAN users who offer high accuracy, substantial health data, and lower bidding costs for training FL model. On the other hand, WBAN users seek to participate in FL model training that offers higher payment and lower associated costs, thereby maximizing their own utility. Additionally, in practice, some WBAN users may only engage in FL model training for a single global iteration due to factors such as user mobility, battery levels, or personal schedules [21, 79]. However, with sufficient local iterations or high local accuracy, the global model can still achieve high accuracy even with just one global iteration[10] [79, 154]. Therefore, the primary goal of this chapter is to maximize overall

[10]We assume that WBAN users involved in FL model training can maintain their connection throughout the duration of a single global iteration.

utility within a single global iteration while adhering to the desired auction properties. Thus, we formulate the optimization problem as the maximization of system utility, which is the linear combination of HSPs' and WBAN users' utilities, as follows:

$$\mathbf{P:} \max_{\mathrm{m}_{p,j}} \sum_{p\in\mathbb{P}} \sum_{j\in\mathbb{J}} \mathrm{m}_{p,j} (\Re_{p,j} + \mathcal{U}_{p,j}) \tag{7.20}$$

Subject to constraints:

$$\sum_{j\in\mathbb{J}} \mathrm{m}_{p,j} \leq 1, \tag{7.20a}$$

$$\mathcal{T}_{p,j} \leq \mathcal{T}_j^{max}, \ \forall \mathrm{m}_{p,j} = 1, \tag{7.20b}$$

$$\epsilon_j^{min} \leq \epsilon_{p,j} \leq \epsilon_j^{max}, \ \forall \mathrm{m}_{p,j} = 1, \tag{7.20c}$$

$$\text{Eqs. } (7.17) - (7.19), \tag{7.20d}$$

$$\mathrm{m}_{p,j} \in \{0, 1\}, \tag{7.20e}$$

$\forall p \in \mathbb{P}, \forall j \in \mathbb{J}$. Constraint (7.20a) restricts each WBAN user to participating in the training of only one FL model. Constraint (7.20b) ensures that the selected WBAN users meet the FL training deadline[11]. Constraint (7.20c) ensures that the privacy budget is appropriately balanced. Constraint (7.20d) guarantees that the auction mechanism adheres to the properties specified in Eqs. (7.17)-(7.19). Lastly, constraint (7.20e) represents the binary decision variable.

The formulated optimization problem **P** is a 0-1 Integer Programming (IP) problem with binary decision variable $\mathrm{m}_{p,j}$, which is NP-hard since its feasibility problem is strongly NP-complete [25, 98]. Due to the high complexity and conditional nature of the formulated optimization problem **P**, this chapter introduces an auction-based suboptimal solution for the maximization problem.

[11] In a practical FL scenario, the maximum FL time T_j^{max} in constraint Eq. (7.20b) is typically determined by the coordinating entity responsible for managing the FL process or by the model owners [26, 155].

7.3 Proposed Solution

This section presents the proposed ADFL algorithm and its theoretical analysis.

7.3.1 ADFL Algorithm

To achieve better global FL model performance within a limited budget $\mathcal{B}_j$, HSPs prioritize selecting WBAN users with higher valuations and lower bidding costs. Unlike conventional auction mechanisms that primarily rely on bidding costs, the FL auction mechanism must consider multiple attributes, including costs (data collection, privacy, computation, and transmission costs) and valuations (reputation, model accuracy, and amount of health data samples used in model training). This comprehensive approach is crucial for selecting WBAN users and designing a more practical and fair incentive mechanism. It ensures that WBAN users with high reputation, model accuracy, large datasets, and lower costs are more likely to be selected and receive incentives accordingly. Thus, we define the *cost per unit valuation* of WBAN user p participating in training of FL model j as follows:

$$\vartheta_{p,j} = \frac{\hat{\mathbb{C}}_{p,j}}{\mathbb{W}_{p,j}}. \tag{7.21}$$

Proposed Algorithm 7.1 outlines the procedure for selecting WBAN users for each FL model and calculating their respective payments. The algorithm begins by initializing binary variables and payments to zero (line 1). It then creates a list of candidate FL models, $\mathscr{N}_p$, for each WBAN user, sorted in non-decreasing order of their cost per unit valuation $\vartheta_{p,j}$ (line 2). Following this, a list of WBAN users who satisfy FL training deadline and privacy budget constraints is constructed as $\mathscr{M}_j = \{p : \mathcal{T}_{p,j} \leq \mathcal{T}_j^{max}$ and $\epsilon_j^{min} \leq \epsilon_{p,j} \leq \epsilon_j^{max}\}$ for each FL model (line 3). In each iteration, the algorithm initializes the candidate list for FL model j to an empty set, i.e., $\mathscr{M}_j' = \varnothing$ (line 5). For each WBAN user p with a non-empty candidate list

Algorithm 7.1: ADFL Algorithm

Input: $\vartheta_{p,j}, \mathbb{W}_{p,j}, \hat{\mathbb{C}}_{p,j}, \epsilon_j^{min}, \epsilon_j^{max}, \mathcal{T}_j^{max}, \mathcal{B}_j, \forall p \in \mathbb{P}, \forall j \in \mathbb{J}$
Output: $\mathrm{m}_{p,j}, pay_{p,j}, \forall p \in \mathbb{P}, \forall j \in \mathbb{J}$

```
1  Initialize m_{p,j} = 0 and pay_{p,j} = 0, ∀p ∈ ℙ, ∀j ∈ 𝕁
2  Create a list of candidate FL models, denoted as 𝒩_p, in non-decreasing order of cost
   per unit valuation ϑ_{p,j}
3  Create candidate list of WBAN users for each FL model:
   ℳ_j = {p : 𝒯_{p,j} ≤ 𝒯_j^{max} and ϵ_j^{min} ≤ ϵ_{p,j} ≤ ϵ_j^{max}}
4  while ∃n : 𝒩_p ≠ ∅ do
5      Set candidate list ℳ'_j = ∅, ∀j ∈ 𝕁
6      for every WBAN user p with 𝒩_p ≠ ∅ do
7          Add WBAN user p to the candidate list of first FL model j in 𝒩_p if p ∈ ℳ_j,
           i.e., ℳ'_j ← ℳ'_j ∪ {p}
8          Remove j from 𝒩_p, i.e., 𝒩_p ← 𝒩_p \ {j}
9      for every FL model j ∈ 𝕁 do
10         Sort WBAN users in ℳ'_j in non-decreasing order of valuation, i.e.,
           ϑ_{1,j} ≤ ϑ_{2,j} ≤ ··· ≤ ϑ_{p',j}
11         Find the largest index 𝒦 in ℳ'_j that satisfies condition: ϑ_{𝒦,j} Σ_{p≤𝒦} 𝕎_{p,j} ≤ ℬ_j
12         Add all WBAN users from index 1 to 𝒦 to the selected set 𝒮_j and set binary
           variable m_{p,j} = 1
13         for every WBAN user p ∈ 𝒮_j do
14             Set payment of WBAN user p for joining FL model j, i.e.,
               pay_{p,j} = 𝕎_{p,j} ϑ_{𝒦,j}
15             Set the candidate list 𝒩_p = ∅
16         ℬ_j ← ℬ_j − Σ_{p∈𝒮_j} pay_{p,j}
```

($\mathscr{N}_p \neq \varnothing$), the algorithm adds p to the candidate list of the first FL model j in $\mathscr{N}_p$ if $p \in \mathscr{M}_j$ (line 7), and then removes j from $\mathscr{N}_p$ (line 8). Subsequently, the algorithm sorts the WBAN users in $\mathscr{M}'_j$ in non-decreasing order of cost per unit valuation, $\vartheta_{p,j}$ (line 10). It then determines the largest index $\mathcal{K}$ in $\mathscr{M}'_j$ that satisfies the budget constraint: $\vartheta_{\mathcal{K},j} \sum_{p \leq \mathcal{K}} \mathbb{W}_{p,j} \leq \mathcal{B}_j$ (line 11), adds all WBAN users from index 1 to $\mathcal{K}$ to the selected set, and updates the binary variables (line 12). Here, $\vartheta_{\mathcal{K},j}$ represents the valuation of the $\mathcal{K}^{th}$ WBAN user in $\mathscr{M}'_j$. Furthermore, the algorithm determines the payment for all WBAN users in the selected set and then clears the candidate list for these users (lines 13-15). Finally, it updates the budget for the FL model (line 16). This iterative process continues until all candidate lists of WBAN users are empty (lines 4-16). Thus, the system utility can be calculated based on $\mathrm{m}_{p,j}$ and $pay_{p,j}$.

7.3.2 Theoretical Analysis of the Proposed ADFL Algorithm

In this section, we present the theoretical analysis of the proposed ADFL algorithm.

Theorem 7.1 *ADFL algorithm satisfies individual rationality.*

Proof: If a WBAN user, with a bidding cost $\hat{\mathbb{C}}_{p,j}$, is not selected, its utility is zero (Eq. (7.14)). In contrast, if a WBAN user is chosen to join the training of FL model j, the payment is determined as $pay_{p,j} = \mathbb{W}_{p,j}\vartheta_{\mathcal{K},j}$, which is $\mathbb{W}_{p,j}(\hat{\mathbb{C}}_{\mathcal{K},j}/\upsilon_{\mathcal{K},j})$. Since the WBAN user with a bidding cost $\hat{\mathbb{C}}_{p,j}$ is selected, the following condition holds: $\hat{\mathbb{C}}_{p,j}/\mathbb{W}_{p,j} \leq \hat{\mathbb{C}}_{\mathcal{K},j}/\upsilon_{\mathcal{K},j}$. This implies $\hat{\mathbb{C}}_{p,j} \leq \mathbb{W}_{p,j}(\hat{\mathbb{C}}_{\mathcal{K},j}/\upsilon_{\mathcal{K},j})$, i.e., $\hat{\mathbb{C}}_{p,j} \leq pay_{p,j}$. Therefore, the selected WBAN user obtains non-negative utility, i.e., $\Re_{p,j} \geq 0$ (based on Eq. (7.14)). Thus, the proposed ADFL algorithm satisfies individual rationality. $\square$

Definition 7.1 An auction mechanism satisfies incentive compatibility (truthfulness) if and only if [25, 75]:

- *Winner determination rule is monotonous:* If a WBAN user wins the auction with a bidding cost $\hat{\mathbb{C}}_{p,j}$, it can also win the auction by bidding a lower cost $\mathbb{C}'_{p,j} < \hat{\mathbb{C}}_{p,j}$.
- *Payment to WBAN user is the critical value:* The payment to the WBAN user is the maximum possible bidding cost that guarantees the user wins the auction.

Theorem 7.2 *ADFL algorithm satisfies incentive compatibility.*

Proof: To prove incentive compatibility, we need to demonstrate that the auction mechanism follows Definition 7.1.

WBAN user selection rule is monotone: If a WBAN user submits a lower bidding cost $\mathbb{C}'_{p,j} < \hat{\mathbb{C}}_{p,j}$ while other WBAN users' bidding cost remain unchanged, it implies

$\mathbb{C}'_{p,j}/\mathbb{W}_{p,j} < \hat{\mathbb{C}}_{p,j}/\mathbb{W}_{p,j}$. In such a case, the WBAN user will either retain its original position or move up to a higher position in the candidate list, securing its selection, as the WBAN user with the lowest $\mathbb{C}_{p,j}/\mathbb{W}_{p,j}$ is chosen first. Therefore, the WBAN user selection rule is monotone.

Payment to WBAN user is the critical value: Assuming the top $\mathcal{K}$ WBAN users are on the candidate list for FL model j, denoted as $\mathscr{M}'_j$, and the payment to the selected WBAN user is $pay_{p,j}$. If the bidding cost of the selected WBAN user, $\hat{\mathbb{C}}_{p,j}$, is less than the payment, i.e., $\hat{\mathbb{C}}_{p,j} < pay_{p,j} = \mathbb{W}_{p,j}(\hat{\mathbb{C}}_{\mathcal{K},j}/\upsilon_{\mathcal{K},j})$, meaning $\hat{\mathbb{C}}_{p,j}/\mathbb{W}_{p,j} < \hat{\mathbb{C}}_{\mathcal{K},j}/\upsilon_{\mathcal{K},j}$, then WBAN user p will remain among the top $\mathcal{K}$ users on the candidate list. Consequently, since $\sum_{p\in\mathbb{P}} \mathbb{W}_{p,j}(\hat{\mathbb{C}}_{\mathcal{K},j}/\upsilon_{\mathcal{K},j}) \leq \mathcal{B}_j$, only the top $\mathcal{K}$ WBAN users, including p, will be selected. However, if the bidding cost $\hat{\mathbb{C}}_{p,j}$ exceeds the payment, i.e., $\hat{\mathbb{C}}_{p,j}/\mathbb{W}_{p,j} > \hat{\mathbb{C}}_{\mathcal{K},j}/\upsilon_{\mathcal{K},j}$, WBAN user p will fall below the top $\mathcal{K}$ users. As a result, since $\sum_{p\in\mathbb{P}} \mathbb{W}_{p,j}(\hat{\mathbb{C}}_{\mathcal{K},j}/\upsilon_{\mathcal{K},j}) > \mathcal{B}_j$, WBAN user p will not be selected. Therefore, payment to the selected WBAN user represents the critical value.

Thus, ADFL algorithm satisfies incentive compatibility. □

Theorem 7.3 *ADFL algorithm satisfies budget feasibility.*

Proof: In the proposed auction, a WBAN user who is not selected for FL model training receives zero payment. Conversely, if selected, the payment is determined as $pay_{p,j} = \mathbb{W}_{p,j}\vartheta_{\mathcal{K},j}$, i.e., $\mathbb{W}_{p,j}(\hat{\mathbb{C}}_{\mathcal{K},j}/\upsilon_{\mathcal{K},j})$. Furthermore, the algorithm ensures that the top $\mathcal{K}$ WBAN users are chosen in each iteration while adhering to the budget constraint for each model, i.e., $\vartheta_{\mathcal{K},j} \sum_{p\leq\mathcal{K}} \mathbb{W}_{p,j} \leq \mathcal{B}_j$ (lines 11-12). Thus, the ADFL algorithm satisfies budget feasibility. □

Theorem 7.4 *Computational time complexity of the ADFL algorithm is $O(JP^2 \log P)$.*

Proof: In Algorithm 7.1, the initialization of variables in line 1 has a time complexity of $O(PJ)$. Creating candidate lists for both FL models and WBAN users (lines 2-3) takes $O(PJ \log P)$, as the number of WBAN users typically exceeds the number of FL models. Initializing $\mathscr{M}_j'$ for all FL models in line 5 has a maximum time complexity of $O(J)$. Lines 6-8 have a time complexity of $O(P)$. Sorting all WBAN users in line 10 takes $O(P \log P)$. Additionally, lines 11 and 12 each take $O(P)$. The for loop in lines 13-15 iterates P times, and updating the budget in line 16 takes constant time, i.e., $O(1)$. As lines 10-16 are executed for each FL model, i.e., J times, the time complexity of lines 9-16 is $O(JP \log P)$. Since the number of iterations depends on the number of WBAN users, i.e., P, the time complexity for lines 4-16 is $O(P(J+P+JP\log P))$, i.e., $O(JP^2 \log P)$. Thus, overall time complexity of Algorithm 7.1 is $O(PJ + PJ\log P + JP^2 \log P)$, i.e., $O(JP^2 \log P)$, indicating the computational efficiency of the proposed ADFL algorithm. $\square$

Theorem 7.5 *ADFL algorithm gives correct solution to the formulated problem* $\boldsymbol{P}$*.*

Proof: When the ADFL algorithm terminates, the while loop (lines 4-16) ensures that constraint (7.20a) is satisfied. Specifically, the candidate list of WBAN user p is emptied after selection in line 15, ensuring that WBAN user p is chosen for FL model j only. Additionally, constraints (7.20b)-(7.20c) are not violated, as WBAN user p is selected for FL model j only if the user is in the candidate list of FL model j, which includes WBAN users meeting FL training deadline and privacy budget constraints (line 3). Based on Theorems 7.1-7.3, the ADFL algorithm satisfies constraint (7.20d). Furthermore, constraint (7.20e) is satisfied, as $\mathrm{m}_{p,j}$ is initialized to 0 in line 1 and updated to 1 only when WBAN user p is selected in line 12. Thus, the proposed ADFL algorithm provides a feasible solution to the formulated problem $\mathbf{P}$. By combining Theorem 7.4 with the preceding arguments, we can affirm that ADFL algorithm provides a feasible solution to the formulated problem $\mathbf{P}$ in polynomial time. Therefore, the correctness of the proposed ADFL algorithm is guaranteed. $\square$

Table 7.1: Parameter setting

Parameter	Value	Parameter	Value
P	$[1000, 2400]$ [71]	ε_p [71]	$[1, 2.5] \times 10^{-28}$
J	$[10, 100]$	$\pi_{p,j}$ [70]	$[10^{-5}, 10^{-4}]$
$\mathcal{T}_j^{max}$, $\mathrm{I}_{p,j}$ [72]	1 s, $[1, 1000]$	$\mathscr{U}_j$ [152], $\mho_{p,j}$ [76]	$[4, 7]$ Mb, $[0.2, 1]$
$\varsigma_{p,j}$ [153], $\mathcal{B}_j$	1, $[1500, 2000]$	$\mathscr{r}_{p,j}$ [152]	$[20, 30]$ Mb/s
$\epsilon_{p,j}$ [73], $\mathrm{f}_{p,j}$	$[0, 25]$, 0.001	ϖ_p [82], $\top$	10 W, 0.04
$\epsilon_j^{min}, \epsilon_j^{max}$	$5, 20$	$\mathscr{Y}$	$[50, 150]$
$\zeta_{p,j}$ [70]	$[10^{-5}, 10^{-4}]$	$\mathscr{w}_1, \mathscr{w}_2$	0.5, 0.5
$\xi_{p,j}$ [73]	$[1.4, 3]$	$\mathfrak{w}_p$ [70]	$[10^{-2}, 10^{-1}]$
$\mathrm{g}_{p,j}$ [79]	$[10, 20]$ MHz	$\mathscr{w}_j$ [82]	$[50, 90]$ cycles

7.4 Performance Evaluation

This section demonstrates the effectiveness of the proposed ADFL algorithm through extensive experimentation conducted on a Windows 10 Home PC equipped with an Intel® Core™ i7-10750H @ 2.60 GHz processor and 16 GB of memory, using Python 3. The simulation results are presented in Section 7.4.1, and the results on real-world data, using UCI HAR [156] and MIT-BIH Arrhythmia Detection [157] datasets, are discussed in Section 7.4.2.

We simulate a multiple FL system, considering scenarios with 1000 to 2400 WBAN users and 10 to 100 FL models from different HSPs to be trained simultaneously [71], with default settings of 2000 WBAN users and 20 FL models [73]. The number of health data samples provided by each WBAN user for FL models is randomly selected from the range [1, 1000] [72], and the budget for each FL model is fixed at 1500 unless otherwise specified. Privacy budgets for WBAN users participating in the FL training are chosen from the range [0, 25] [73]. The size of FL models varies between $[4, 7]$ Mb [152], while the data transfer rate between WBAN users and HSPs is set between $[20, 30]$ Mb/s [152]. The computation capacity (CPU frequency) of WBAN users is uniformly distributed between $[10, 20]$ MHz [79]. The required CPU cycles to train one health data sample at the WBAN for each FL model is chosen between $[50, 90]$ cycles [82]. Further details on the parameters used in the experiments are provided in

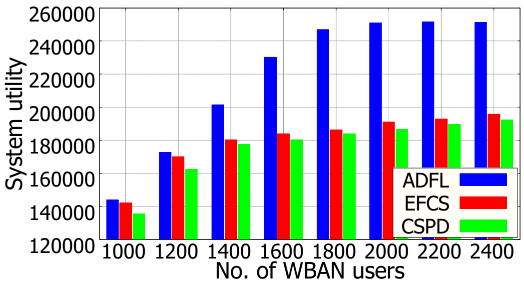
Table 7.1.

We compare our results with two benchmark schemes: Envy-Free Client Selection (EFCS) [72] and Client Selection and Payment Determination (CSPD) [75]. These existing works, EFCS and CSPD, closely resemble our proposed work as they address client selection in differentially private FL settings and incorporate incentive mechanisms through auction methods. In EFCS, after collecting bidding costs from each client, unit costs for each data sample are calculated and sorted in ascending order to determine winners and their payments. In contrast, CSPD computes each client's quality per unit cost and sorts them in descending order for winner determination and payment allocation. For a fair comparison, we consider the privacy budget as a measure to evaluate the quality of the client's model in CSPD, as it directly influences model quality [73]. Additionally, in our comparison, we treat the clients used in EFCS and CSPD as equivalent to WBAN users.

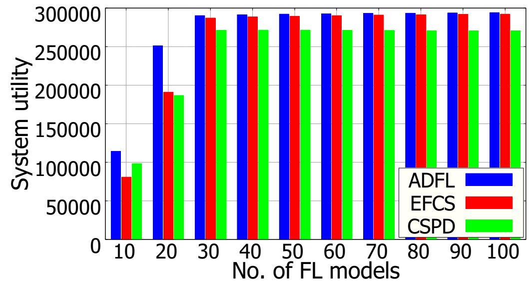
7.4.1 Simulation Results

In this section, we evaluate the effectiveness of the proposed ADFL algorithm through comprehensive analysis, particularly focusing on utility variations, number of selected WBAN users, total valuation, and total data involved across various settings, as discussed below.

System utility: Fig. 7.3a compares the system utility of the proposed ADFL algorithm with that of EFCS and CSPD as the number of WBAN users varies from 1000 to 2400. We observe that system utility increases with the number of WBAN users, and the proposed ADFL algorithm outperforms EFCS and CSPD, achieving, on average, 15.9% and 18.08% higher utility, respectively. This improvement is due to the ADFL algorithm's incorporation of valuation, which accounts for accuracy, reputation, and number of health data samples, whereas EFCS and CSPD rely solely on the number of health data samples and the local model quality, respectively. Moreover, the sys-



(a) System utility vs WBAN users



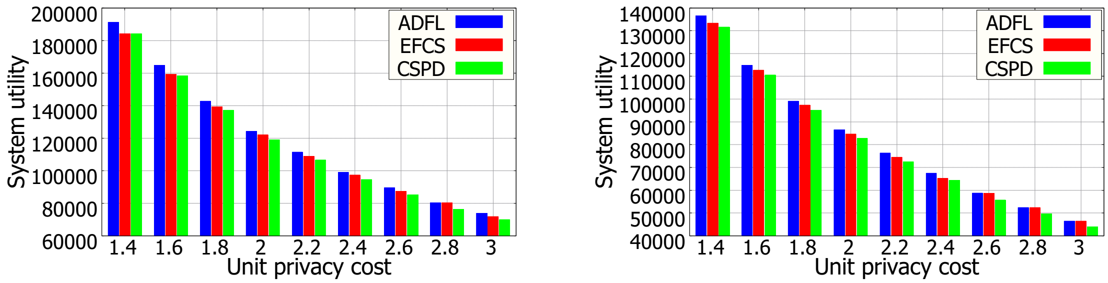
(b) System utility vs FL models

Fig. 7.3. System utility vs WBAN users and FL models.

tem utility initially increases and subsequently becomes constant, as additional WBAN users cannot be selected without violating the constraints of the formulated problem **P** due to the fixed budget for FL models.

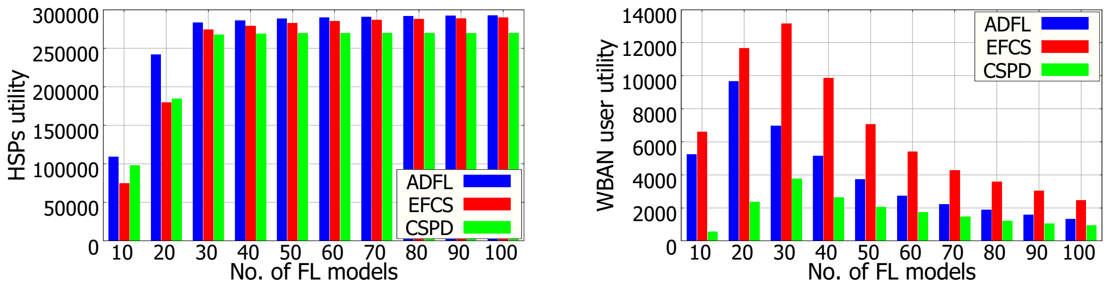
Fig. 7.3b compares the system utility of the proposed ADFL algorithm with that of EFCS and CSPD as the number of FL models varies from 10 to 100. We observe from the results that system utility increases with the number of FL models, and ADFL outperforms both EFCS and CSPD. This is because ADFL considers the valuation of WBAN users, incorporating factors such as accuracy, reputation, and number of health data samples, whereas EFCS and CSPD rely solely on the number of health data samples and the local model quality, respectively, to determine the winning WBAN users and their payments. Additionally, the results indicate that system utility initially rises and then becomes constant, indicating a saturation point where no more WBAN users can be selected.

Fig. 7.4a presents the comparison of system utility when the privacy budget of WBAN users fixed at 15, while Fig. 7.4b shows the comparison when the privacy budget fixed at 20, with unit privacy costs varying from 1.4 to 3.0. We observe from the results that system utility decreases as the unit privacy cost increases. This is because a higher unit privacy cost raises the overall cost for WBAN users, resulting in the selection of fewer users and a subsequent reduction in system utility. Additionally,



(a) Privacy budget $(\epsilon_{p,j}) = 15$ **(b)** Privacy budget $(\epsilon_{p,j}) = 20$

Fig. 7.4. System utility vs unit privacy cost $(\xi_{p,j})$.

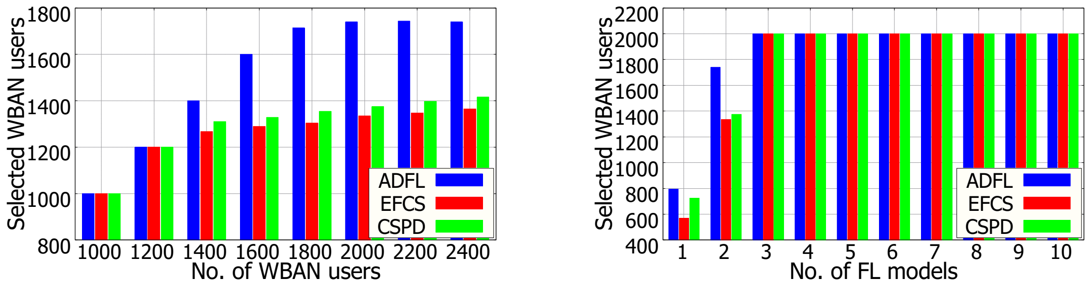


(a) HSPs' utility **(b)** WBAN users' utility

Fig. 7.5. Utility of HSPs and WBAN users on FL models.

comparing Figs. 7.4a and 7.4b shows that system utility is lower with a higher privacy budget. Specifically, for a given unit privacy cost, system utility is reduced when the privacy budget is set at 20 compared to 15, as the higher privacy budget increases the cost for WBAN users, leading to the selection of fewer users and a decrease in system utility.

Utility of HSPs and WBAN users: Fig. 7.5a compares the utility of HSPs as the number of FL models ranges from 10 to 100. We observe from the result that the utility of HSPs rises with the number of FL models, as more WBAN users can be selected due to the increased total budget in the system. However, the utility becomes constant when selecting further WBAN users becomes infeasible due due to their limited availability. Additionally, the proposed ADFL outperforms EFCS and CSPD because it selects more number of WBAN users, thereby increasing the total valuation for HSPs and resulting



(a) Selected vs WBAN users **(b)** Selected vs FL models

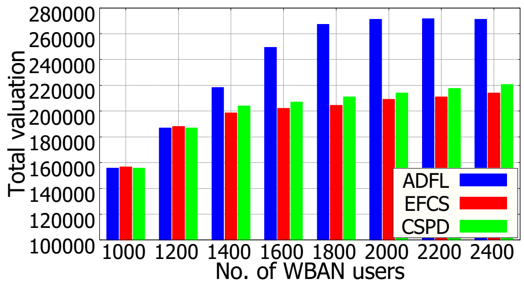
Fig. 7.6. Total selected WBAN users.

in higher utility.

Fig. 7.5b compares the utility of WBAN users as the number of FL models varies from 10 to 100. We observe from the result that the WBAN users' utility increases up to a certain point and then gradually decreases as the number of FL models continues to rise. The reason is that the payment to WBAN users initially increases but eventually decreases as the number of WBAN users that can be selected per FL model diminishes. Additionally, the utility of WBAN users obtained by EFCS is higher than that achieved by the proposed ADFL. However, the system utility obtained by EFCS is lower compared to that of the proposed ADFL (see Fig. 7.3).

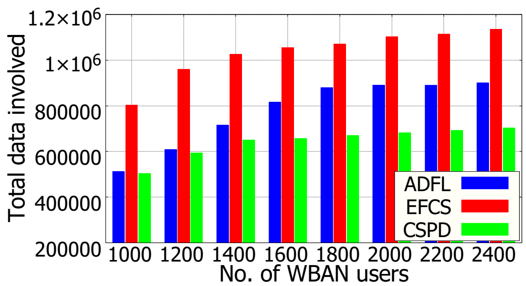
Number of selected WBAN users: Fig. 7.6a illustrates how the number of selected WBAN users varies with the total number of WBAN users in the system. The result shows that as the total number of WBAN users increases, the number of selected users also rises. However, a saturation point is reached at 1800 WBAN users, beyond which adding more WBAN users violates the constraints in problem $\mathbf{P}$ due to the fixed budget of FL models. A similar trend is observed in Fig. 7.6b, where the number of selected WBAN users grows with the number of FL models and then becomes constant, indicating that no additional WBAN users are available for selection.

Total valuation: Fig. 7.7 shows the change in total valuation as the number of WBAN users and FL models in the system varies. The results in Figs. 7.7a and



(a) Valuation vs WBAN users **(b)** Valuation vs FL models

Fig. 7.7. Total valuation of FL models.

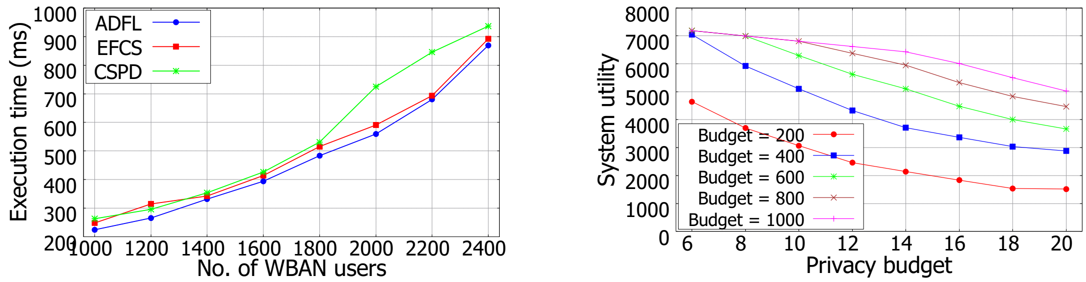


(a) Total data vs WBAN users **(b)** Total data vs FL models

Fig. 7.8. Total data involved in FL models.

7.7b show an increase in total valuation with the rising number of WBAN users and FL models. This is because a larger number of WBAN users with higher valuations can be selected as both the number of WBAN users and FL models increase (see Fig. 7.6). Moreover, the proposed ADFL algorithm outperforms EFCS and CSPD because it considers valuation in the WBAN user selection process, enabling the selection of WBAN users with higher valuations. However, the total valuation becomes constant after reaching a certain point due to the fixed budget of FL model or the fixed number of available WBAN users.

Total data involved: Fig. 7.8 illustrates the total health data provided by selected WBAN users as the number of WBAN users (Fig. 7.8a) and the number of FL models (Fig. 7.8b) in the system vary. The results show an increase in the total health data provided by the selected WBAN users as the number of WBAN users and FL models



(a) Execution time. **(b)** System utility vs privacy budget.

Fig. 7.9. Execution time and system utility vs privacy budget.

increase. Moreover, the total health data involved in EFCS exceeds that of the proposed ADFL algorithm. This is because EFCS selects WBAN users and determines their payments based solely on the amount of health data provided, without considering other factors such as valuation and model accuracy. However, the system utility and the number of selected WBAN users achieved by EFCS are lower compared to those of the proposed ADFL algorithm (see Figs. 7.3 and 7.6).

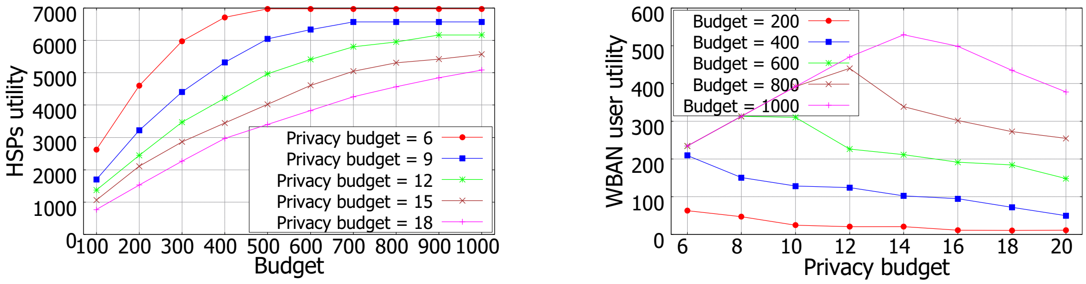
Execution time: Fig. 7.9a compares the execution time of ADFL with that of EFCS and CSPD in various scenarios. We observe from the result that the execution time increases as the number of WBANs increases. This is because, with more WBAN users, the total possibilities for selecting users to participate in each FL model also increase, leading to higher execution times. Additionally, the execution times of EFCS and CSPD are almost the same as that of the proposed ADFL algorithm. However, the system utility obtained by EFCS and CSPD is lower than that of the proposed ADFL algorithm.

7.4.2 Results on Real-World Data

In this section, we conduct a performance comparison of the proposed ADFL algorithm using a real-world dataset, considering two FL models and 50 WBAN users across different settings. Our experimental setup utilizes the UCI Human Activity Recognition (HAR) and MIT-BIH arrhythmia detection datasets for training the FL models, following the data distribution settings specified in [77]. The UCI HAR dataset consists of 10299 data samples collected from 30 smartphone users, which are divided into 9000 training samples and 1299 test samples. These training samples are distributed among 50 WBAN users, with 10 users receiving 360 samples each, 25 users receiving 180 samples each, and 15 users receiving 60 samples each. In contrast, the MIT-BIH arrhythmia dataset contains 26490 data samples, with 18000 allocated for training and 8490 for testing. These training samples are also distributed among 50 WBAN users, with 10 users receiving 720 samples each, 25 users receiving 360 samples each, and 15 users receiving 120 samples each. Our experiment assumes that the health datasets are available to each WBAN user for training FL models.

Fig. 7.9b analyzes five scenarios with fixed budgets for each FL model set at 200, 400, 600, 800, and 1000, while varying the privacy budget for WBAN users from 6 to 20. We observe from the result that the system utility decreases as the privacy budget increases due to the rising privacy costs for WBAN users. This leads to higher costs for WBAN users and, consequently, lower utilities for both the users and the system. Additionally, we observe that system utility is higher with a larger budget, as it allows for the selection of more WBAN users. This, in turn, increases the utilities of both WBAN users and HSPs, resulting in a higher overall system utility.

Fig. 7.10a considers five scenarios where the privacy budget for WBAN users is fixed at 6, 9, 12, 15, and 18, while the budget for each FL model ranges from 100 to 1000. The result shows that HSPs' utility increases with the FL model budget, as a larger budget allows for the selection of more WBAN users, thereby increasing total valuation and utility. However, beyond a certain point, further increases in the FL model budget have no impact on utility due to the limited number of available WBAN users. Additionally, HSPs' utility is lower when the privacy budget is higher because the increased cost of WBAN users reduces the number of users selected and overall



(a) HSPs' utility vs budget **(b)** WBAN utility vs budget

Fig. 7.10. Comparison of HSPs' and WBAN users' utility.

valuation, leading to lower utility.

Fig. 7.10b considers five scenarios where FL model budgets are fixed at 200, 400, 600, 800, and 1000, while WBAN users' privacy budgets range from 6 to 20. We observe that as the privacy budget of WBAN users increases, their utility also rises. However, as privacy budgets increase further, utility begins to decrease due to the higher costs incurred by WBAN users, while the total budget remains constant, resulting in lower payments and reduced utility. Additionally, for a given privacy budget, utility is higher when the budget for the FL model is larger. For instance, when budgets are set at 600, 800, and 1000, utility increases with higher privacy budgets because more WBAN users can be selected. In contrast, when the budget is smaller (e.g., 200 or 400), increasing the privacy budget leads to lower utility for WBAN users.

7.5 Summary

This chapter proposed an auction-based incentive mechanism and WBAN users selection framework to enable the parallel training of multiple FL models using heterogeneous WBAN users' health data, ensuring privacy of data. Additionally, an optimization problem is formulated to maximize system utility, which is NP-hard, and incorporated a cost model that includes data collection, computation, communication, and privacy. The proposed auction-based algorithm integrates factors such as local model accuracy, user reputation, and data volume to solve the formulated problem efficiently. Furthermore, theoretical analyses of auction properties, including individual rationality, incentive compatibility, budget feasibility, and computational efficiency, are provided. Extensive simulations and real-world data analysis demonstrated the efficacy of the proposed algorithm, achieving average utility improvements of 15.9% and 18.08% compared to state-of-the-art methods.

In the next chapter, we summarize the key conclusions of this thesis and presents future research directions.