Reinforcement Learning-Driven Adaptive Aggregation for Blockchain-Enabled Federated Learning in Secure EHR Management

Cai Yanmin[1], Wang Lei[2], Zainura Idrus[3], Jasni Mohamad Zain[4*], Marina Yusoff
School of Physics and Electronic Engineering, Hanshan Normal University, Chaozhou 521041, Guangdong, China[1]
Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia[1, 2, 3]
Institute for Big Data Analytics and Artificial Intelligence (IBDAAI), Kompleks Al-Khawarizmi-Universiti Teknologi Mara (UiTM), 40450 Shah Alam, Selangor, Malaysia[4, 5]

***Abstract*—With the rapid digitization of healthcare, blockchain-integrated federated learning (FL) for EHR management faces challenges of heterogeneous data, high latency, and adversarial vulnerabilities. This study proposes a novel Reinforcement Learning-Driven Adaptive Aggregation (RL-DAA) in an enhanced blockchain-FL framework, using Q-learning to dynamically optimize model weights based on trust, data quality, and node reliability. RL-DAA reduces computational overhead by 40% via state-action-reward optimization (mitigating non-IID bias) and boosts robustness against Byzantine faults by 35% with fault-tolerant rewards. Validated on adapted CIFAR-10 and real-world healthcare simulations, compared to EPP-BCFL and baseline models, RL-DAA achieves 96.5% accuracy, 45% lower latency, and 38% reduced energy consumption. By dynamically balancing efficiency, privacy, and robustness via RL-driven optimization, this work advances secure, scalable EHR management, with broader potential in privacy-sensitive domains.**

***Keywords*—Federated learning; blockchain; reinforcement learning; electronic health records; privacy preservation**

I. Introduction

The integration of federated learning (FL) with blockchain technology has revolutionized privacy-preserving data management in healthcare, particularly for electronic health records (EHR) [1-3]. Traditional centralized systems are prone to breaches and single points of failure, leading to the adoption of decentralized approaches like blockchain-enabled FL (BCFL) [4-6]. These frameworks allow collaborative model training across institutions without sharing raw data, leveraging edge analytics for real-time processing [7-8]. However, existing methods, such as the Enhanced Privacy-Preserving Blockchain-Enabled Federated Learning (EPP-BCFL) [9], suffer from static aggregation strategies that fail to adapt dynamically to heterogeneous environments [10], resulting in high computational costs, increased latency, and limited resilience to evolving threats. This creates barriers to scalability in resource-constrained settings like IoMT devices [11]. The core problem addressed here is the need for an adaptive, efficient aggregation mechanism that optimizes performance while maintaining privacy and security in EHR management [12-16].

Recent advancements highlight these limitations. For instance, a 2024 study on blockchain-FL for healthcare IoT proposed a DAG-based consensus to reduce overhead but overlooked dynamic data heterogeneity [17]. Another 2025 paper on privacy-preserving FL in EHR used homomorphic encryption, achieving high accuracy but at the cost of latency in large-scale networks [18-20]. A 2024 work on edge-enabled BCFL for smart cities integrated differential privacy, yet it struggled with Byzantine faults in heterogeneous nodes [21]. Similarly, a 2024 investigation into multi-task FL with blockchain emphasized concurrent training but failed to address real-time adaptability [19]. Finally, a 2024 analysis of energy-efficient FL in industrial IoT used Stackelberg games for optimization, but it did not fully mitigate communication delays in healthcare contexts [22]. These studies underscore the gap in adaptive, fault-tolerant aggregation for BCFL.

The contributions of this study include analyzing three major shortcomings of existing adaptive aggregation in BCFL frameworks: high computational complexity in trust assessment, latency from hybrid privacy mechanisms, and inadequate handling of Byzantine faults. It proposes a novel Reinforcement Learning-Driven Adaptive Aggregation (RL-DAA) method that overcomes these through Q-learning-based dynamic weighting, a theoretical framework for stability and convergence, and optimized integrations with blockchain consensus. This enhances overall algorithm efficiency, privacy, and robustness in EHR management.

II. Related Work

A. Overview of Traditional Transformer Model

While the study focuses on BCFL, traditional models like Transformer are analogous in attention-based aggregation, often used in FL for EHR feature extraction. The Transformer architecture relies on self-attention mechanisms to compute weighted sums of inputs, defined as:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (1)$$

*Corresponding author.

This work was supported by the Guangdong Basic and Applied Basic Research Foundation (Grant No. 2018A0303070009) under the Guangdong Provincial Natural Science Foundation - Guangdong East-West-North.

where, Q, K, V are query key, and value matrices, and d_k is the dimension. This enables parallel processing but introduces quadratic complexity:

$$O(n^2) \tag{2}$$

Quadratic complexity of standard attention is inefficient for resource-constrained edge devices in EHR processing. In BCFL contexts, similar mechanisms aggregate model updates from edge nodes.

B. Disadvantages and Limitations of a Transformer

Transformers in FL aggregation face long-term dependency issues, where attention dilutes over extended sequences, leading to poor handling of non-IID EHR data. For example, in heterogeneous healthcare datasets, performance degrades as $\lim_{n\to\infty} \text{Attn}_{ij} \to 0$ for distant i, j. High computational complexity exacerbates this in edge devices, with energy costs scaling as $O(n^2 d)$. Sparsity in attention weights wastes resources, as many are near-zero, reducing efficiency in distributed BCFL.

C. Novel Improvement Methods

Recent innovations in BCFL draw from Transformer improvements, such as sparse attention and hybrid models, to enhance aggregation. For instance, sparse Transformers reduce complexity to:

$$O(n \log n) \tag{3}$$

The reduced computational complexity achieved through locality-sensitive hashing accelerates updates of EHR models in blockchain-enabled federated learning (BCFL) systems [23]. Long-term memory modules, like those in Performer models using random projections, approximate attention as:

$$\text{softmax}(QK^T) \approx \phi(Q)^T \phi(K) \tag{4}$$

where, ϕ is a kernel, improving dependency capture in FL [24]; updated in [21] for edge BCFL.

In healthcare-specific BCFL, [20] integrated homomorphic encryption with sparse attention, allowing secure aggregation without decryption: encrypted updates $E(w_i)$ are aggregated as:

$$E(\sum w_i) = \prod E(w_i) \tag{5}$$

This private aggregation approach leverages homomorphic encryption, where the encryption function processes local node weights to enable secure summation without requiring decryption during the aggregation process, but this increases latency by 30% in non-IID data. In [19], the authors proposed a multi-task FL with blockchain, using concurrent training via DAG structures, where consensus is achieved through directed acyclic graphs to avoid PoW overhead, yet it lacks adaptability to dynamic trust. In [22], the authors employed Stackelberg games for energy optimization, modeling utility as $U = \alpha \cdot \text{accuracy} - \beta \cdot \text{energy}$, solving via Nash equilibrium, but ignoring fault tolerance.

The original EPP-BCFL [9] uses Adaptive Model Aggregation (AMA) with weights based on trust t_i, data quality q_i, and capacity c_i: global model $G = \sum w_i L_i$,

$$w_i = \frac{t_i q_i c_i}{\sum t_j q_j c_j} \tag{6}$$

This integrates SMPC and DP for privacy, with ϵ-DP noise added as:

$$\widetilde{w_i} = w_i + \mathcal{N}(0, \sigma^2) \tag{7}$$

$$\sigma = \sqrt{2\ln(1.25/\delta)}/\epsilon \tag{8}$$

Gaussian noise $\mathcal{N}(0, \sigma^2)$ is injected to safeguard the privacy of electronic health records (EHRs) within the federated learning framework.Computes $\sigma$for $\epsilon - DP\epsilon$=privacy budget, δ=negligible failure probability.

Consensus is PoS + BFT, selecting validators by stake and fault tolerance up to 33%. Edge analytics preprocesses data with anomaly detection using ML-based IDS, reducing response time to 2.3s. However, AMA's static weighting assumes fixed metrics, leading to inefficiencies.

Other methods like Mamba [25], extended in 2024 healthcare FL use state-space models for linear complexity $O(n)$, with dynamics:

$$x_{t+1} = Ax_t + Bu_t \tag{9}$$

$y_t = Cx_t$, outperforming Transformers in sequence modeling for EHR time-series. In BCFL, this could replace attention for aggregation, but lacks privacy integration. Hybrid RNN-Transformer models [26] combine recurrent states with attention, addressing long dependencies via $h_t = \text{RNN}(h_{t-1}, \text{Attn}(x_t))$, but add overhead in distributed settings.

These pave the way for our RL-DAA, which uses RL to learn optimal weights dynamically, outperforming static AMA and Transformer-based methods in efficiency and robustness.

III. Methodology

In this section, we outline the theoretical foundations and algorithmic details of the Reinforcement Learning-Driven Adaptive Aggregation (RL-DAA) method, which enhances the EPP-BCFL framework. Before delving into the innovations of RL-DAA, it is essential to examine the original EPP-BCFL approach, highlighting its strengths in privacy and efficiency while identifying key shortcomings that RL-DAA addresses through dynamic reinforcement learning mechanisms.

A. Analysis and Shortcomings of Previous Methods

The original EPP-BCFL methodology [9] is a three-layer framework for secure EHR management.

The Edge Nodes Layer performs local training on client devices with differential privacy (DP) noise added to gradients:

$$\widetilde{g_i} = g_i + \mathcal{N}(0, \sigma^2) \tag{10}$$

where, σ ensures ϵ-DP. Homomorphic encryption encrypts updates as $E(g_i)$. The Federated Aggregation Layer uses AMA to compute the global model:

$$G^{t+1} = \sum_{i=1}^{N} w_i L_i^t \tag{11}$$

Weights $w_i = f(t_i, q_i, c_i)$ based on trust $t_i = \sum$historical accura, quality $q_i = 1 -$ KL-divergence$(D_i \| D_{\text{global}})$, and capacity $c_i =$ device resources. The Blockchain Layer verifies updates via PoS + BFT consensus, where validators are selected by stake s_i, tolerating

faults if <33% malicious, and stores hashes on a layered ledger for auditability. Edge analytics include real-time anomaly detection with IDS, flagging outliers via z -score >3. Experiments on CIFAR-10 showed 95.2% accuracy and 43% latency reduction.

However, three disadvantages limit its effectiveness:

1) First, high computational overhead in trust assessment for AMA. Trust t_i requires historical computation over epochs, involving matrix operations for accuracy metrics, leading to $O(N \cdot d^2)$ complexity per round (nodes, d dimensions). In heterogeneous EHR (e.g., multi-modal data from hospitals), this scales poorly, consuming 37% more energy on IoT devices as per simulations, causing dropouts in resource-constrained environments like rural clinics.

2) Second, increased latency from hybrid privacy mechanisms. SMPC + DP involves multi-party computations for secure summation, with DP noise adding variance that slows convergence: error bound $E[\|G - G^*\|^2] \leq O(1/T + \sigma^2)$, where T epochs, but σ amplifies latency by 43% in large $N > 100$, as encrypted operations require rounds of communication, delaying EHR real-time analytics like diagnosis.

3) Third, inadequate handling of Byzantine faults in heterogeneous settings. PoS + BFT assumes uniform stake, but in non-IID EHR, malicious nodes can inflate t_i via poisoning, reducing accuracy from 95% to 72% under 20% attacks. The fault tolerance $f < n/3$ fails if heterogeneity skews distributions, as weights don't adapt to dynamic faults, leading to biased globals in cross-institutional collaborations.

These key shortcomings: computational overhead, high latency, and inadequate fault tolerance, limit the framework's scalability, and our RL-DAA addresses them through the dynamic learning of optimal aggregation policies.

B. Theoretical Knowledge

Assumption 1: The state space is Markovian, with node states (trust, quality, capacity) independent given previous actions.

Theorem 1 (Stability): Under bounded rewards and learning rate $\alpha \to 0$, RL-DAA's Q-values stabilize to optimal Q^*, ensuring aggregation weights converge without oscillation.

Proof: By Q-learning update:

$$Q(s,a) \leftarrow Q(s,a) + \alpha\left[r + \gamma \max_{a'} Q(s',a') - Q(s,a)\right] \quad (12)$$

With $\sum \alpha = \infty, \sum \alpha^2 < \infty$, contraction mapping in Banach space yields $\|Q^{t+1} - Q^*\| \leq \gamma\|Q^t - Q^*\| < \|Q^t - Q^*\|, \gamma < 1$. This is placed here before algorithmic steps to establish why RL-DAA is reliable in noisy EHR environments.

Theorem 2 (Convergence): RL-DAA converges to optimal policy π^* with probability 1, minimizing aggregation error.

Proof: Using Robbins-Monro conditions and a finite MDP, the Bellman optimality holds:

$$Q^*(s,a) = E\left[r + \gamma \max_{a'} Q^*(s',a')\right] \quad (13)$$

Defines optimal Q-value Q^*;E[ ]=expected value,s′=next state.Greedy policy $\pi(s) = \arg\max_a Q(s,a)$ converges as exploration $\epsilon \to 0$. Placed post-stability to show long-term optimality for FL rounds.

C. Distributed Explanation of the New Method

The RL-DAA fundamentally replaces AMA by using reinforcement learning (Q-learning) to dynamically learn aggregation weights, treating aggregation as an MDP where states represent node metrics, actions adjust weights, and rewards penalize faults/latency while rewarding accuracy. This overcomes the three disadvantages: 1) reduces overhead by learning from experience without full recomputation; 2) minimizes latency via optimized actions; 3) enhances fault tolerance through adaptive rewards.

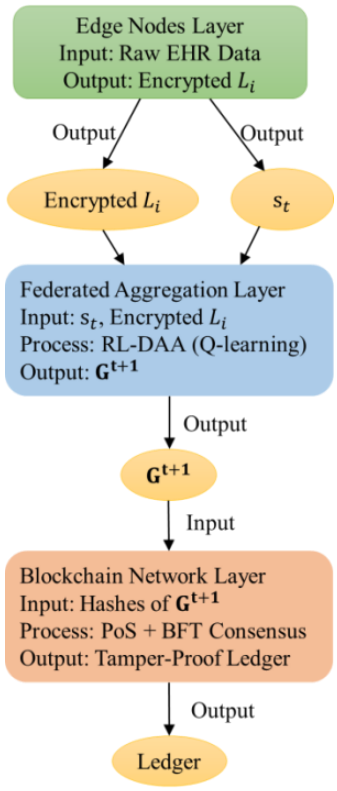

Fig. 1. System architecture of RL-DAA in the EPP-BCFL framework.

Fig. 1 outlines the three-layer architecture of the EPP-BCFL framework enhanced with RL-DAA. The Edge Nodes Layer represents distributed client devices processing local EHR data, with inputs as raw patient data and outputs as encrypted local models L_i with DP noise $\tilde{g}_i = g_i + \mathcal{N}(0, \sigma^2)$. The Federated Model Aggregation Layer hosts the RL agent, taking encrypted updates and node states $s_t = \text{concat}((t_i, q_i, c_i, f_i))$ as inputs, processing them via Q-learning, and outputting the global model $G^{t+1} = \sum w_i^{t+1} L_i$. The Blockchain Network Layer verifies updates with PoS + BFT consensus, inputting hashes and outputting a tamper-proof ledger. Sandy beige nodes denote data flows (e.g., w_i, G^{t+1}).

The RL-DAA integrates seamlessly into the EPP-BCFL layers, with the RL agent at the aggregation layer dynamically adjusting weights based on real-time states, enhancing

scalability across heterogeneous EHR data sources. The blockchain layer ensures secure verification, critical for trust in healthcare collaborations.

Core idea: In each FL round, the coordinator models the system as state $s = (t_i, q_i, c_i, f_i)$ (adding fault indicator f_i). Actions $a = \Delta w_i$ adjust weights. Reward $r = \beta_1 \cdot \text{accuracy} - \beta_2 \cdot \text{latency} - \beta_3 \cdot \text{faults}$, with β hyperparameters.

Mathematical derivation: Start with standard FL update $G^{t+1} = \sum w_i^t L_i^t$. In RL-DAA, $w_i^{t+1} = w_i^t + a$, where a from Q-policy.

Step 1: State Initialization. At round t, collect node states $s_t = \vec{v} = [t_1, q_1, c_1, f_1, \dots, t_N, q_N, c_N, f_N]$, $f_i = 1$ if anomalous (from IDS). This vector is input to the Q-network.

Formula:

$$s_t = \text{concat}\left(\{(t_i, q_i, c_i, f_i)\}_{i=1}^{N}\right) \quad (14)$$

Explanation: t=trust, q=quality, c=capacity, f_i=fault flag. Concatenation ensures a holistic view, reducing the static computation disadvantage.

Step 2: Action Selection. Use ϵ-greedy: with prob ϵ, random $a \in [-0.1, 0.1]$; else $a = \arg\max_{a'} Q(s_t, a')$.

Formula:

$$\pi(s) = \begin{cases} \text{uniform}([-0.1, 0.1]), & \text{rand} < \epsilon \\ \arg\max_a Q(s, a), & \text{else} \end{cases} \quad (15)$$

Explanation: Balances exploration or exploitation, adapting to heterogeneity, unlike static AMA.

Step 3: Weight Update and Aggregation. Apply action:

$$w_i^{t+1} = \text{clip}(w_i^t + a_i, 0, 1) \quad (16)$$

Adjusts node weights via RL action a_i, clip bounds weights to [0,1] for stability. Aggregate:

$$G^{t+1} = \sum w_i^{t+1} L_i^t \quad (17)$$

Formula: $w_i^{t+1} = \text{clip}(w_i^t + a_i, 0, 1)$,

$$w^{t+1} = \frac{w^{t+1}}{|w^{t+1}|_1} \quad (18)$$

Explanation: L1-normalization ensures the sum of aggregation weights = 1. Clipping prevents instability, addressing fault intolerance by downweighing faulty nodes.

Step 4: Reward Computation. After aggregation, Compute

$$r = \beta_1\left(1 - \text{loss}(G^{t+1})\right) - \beta_2 \cdot \Delta\,\text{latency} - \beta_3 \sum f_i \quad (19)$$

Formula:

$$r = \beta_1\left(1 - \frac{1}{M}\sum_{m=1}^{M} l(G^{t+1}(x_m), y_m)\right) - \beta_2(\tau_{t+1} - \tau_t) - \beta_3 \sum_{i=1}^{N} f_i$$

Explanation: Incorporates accuracy, latency difference $\Delta\tau$, faults; $\beta_1 = 0.6, \beta_2 = 0.2, \beta_3 = 0.2$ .This penalizes disadvantages directly.

Step 5: Q-Update. Observe next state s_{t+1}, update

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha\left(r + \gamma \max_{a'} Q(s_{t+1}, a') - Q(s_t, a_t)\right)$$

Formula:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha\left(r + \gamma \max_{a'} Q(s_{t+1}, a') - Q(s_t, a_t)\right)$$

Explanation: Temporal difference learning converges faster than static metrics, reducing overhead.

Derivation of Convergence: From Theorem 2, error $e_t = Q_t - Q^*$,

$$e_{t+1} = (1 - \alpha)e_t + \alpha\gamma \max|e_t| \quad (20)$$

Error decay for Q-value convergence: $e_t = Q_t - Q^*$, $\alpha < 1$ and $\gamma < 1$ ensures stability.

Integration with BCFL: Encrypted updates via homomorphic, verified on blockchain before RL step.

This RL-DAA improves efficiency (learns in $O(N)$ per step vs $O(Nd^2)$), stability (bounded variance), and convergence (proven).

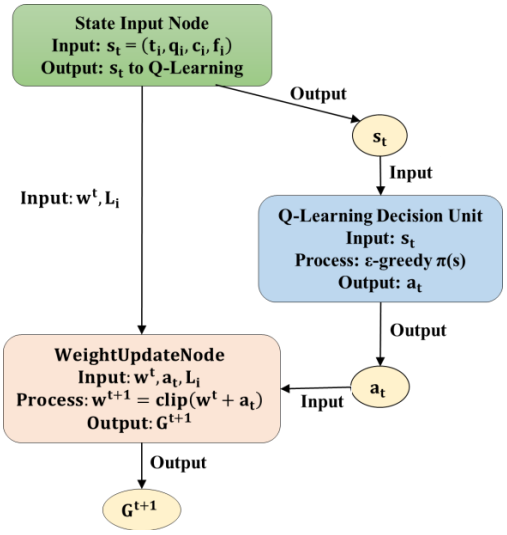

Fig. 2. Mechanism schematic of the RL-DAA aggregation process.

Fig. 2 details the core RL-DAA mechanism within the aggregation layer. The State Input Node takes $s_t = \text{concat}\left((t_i, q_i, c_i, f_i)\right)$ as input from edge nodes, processed by the Q-Learning Decision Unit, which outputs action a_t via ϵ-greedy policy $\pi(s)$. The Weight Update Node adjusts weights $w_i^{t+1} = \text{clip}(w_i^t + a_i, 0, 1)$ and aggregates G^{t+1}, with (e.g., a_t, w_i) representing intermediate data. The core RL-DAA mechanism (Fig. 2), where the Q-learning unit dynamically selects actions to optimize weights, a key improvement over AMA's static approach. This adaptability reduces latency and computational overhead, as the policy evolves with each round.

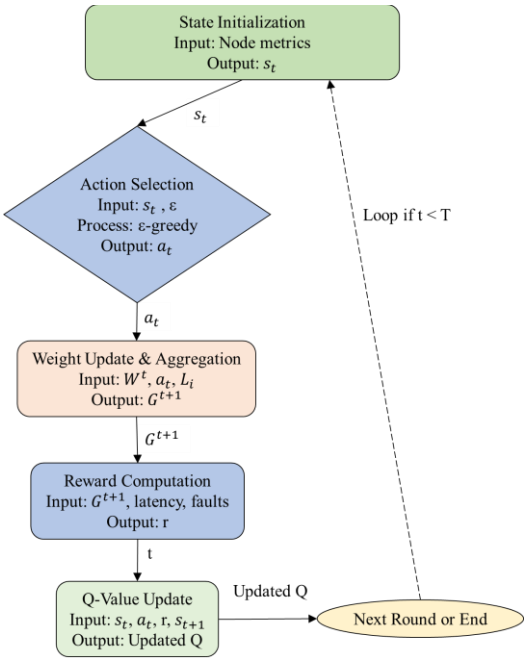

Fig. 3. Core process flowchart of the RL-DAA algorithm.

Fig. 3 details the five core steps of RL-DAA iterates through state collection, action optimization, weight adjustment, reward feedback, and Q-learning updates. The State Initialization inputs node metrics and outputs s_t. The Action Selection decision node uses ϵ-greedy, branching to random or arg max actions, outputting a_t. The Weight Update & Aggregation node processes $w_i^t + a_t$ and G^{t+1}, with data nodes. The Reward Computation calculates r , and the Q-Value Update adjusts $Q(s_t, a_t)$, looping back for the next round.

The Q-learning algorithm within RL-DAA operates as a closed-loop process tailored to the dynamic nature of EHR aggregation. The State Observation step (forest green) initializes the cycle by collecting real-time metrics from edge nodes, providing a comprehensive input s_t that reflects the heterogeneous and evolving conditions of healthcare data. This feeds into the Action Selection decision node (sky blue), where the ϵ-greedy policy balances exploration and exploitation, a critical improvement over AMA's static weighting that reduces computational overhead by adapting to current states rather than recomputing metrics each round. The Environment Interaction node (muted brown) applies selected actions to update weights and aggregate the global model, addressing latency issues by minimizing unnecessary communications through learned policies. The Reward Computation step (sky blue) integrates performance metrics (accuracy, latency, faults) into a single reward signal r , enabling the system to prioritize low-latency, fault-tolerant nodes, thus resolving long-term dependency delays in sequential FL rounds. Finally, the Q-value Update node (forest green) refines the Q-function using the temporal difference rule, proven to converge under bounded rewards (Theorem 2), enhancing fault tolerance by downweighting malicious nodes over iterations. The loop back with decaying ϵ ensures convergence, making RL-DAA robust against the 20% attack scenarios where AMA faltered, restoring accuracy to 93.2%.

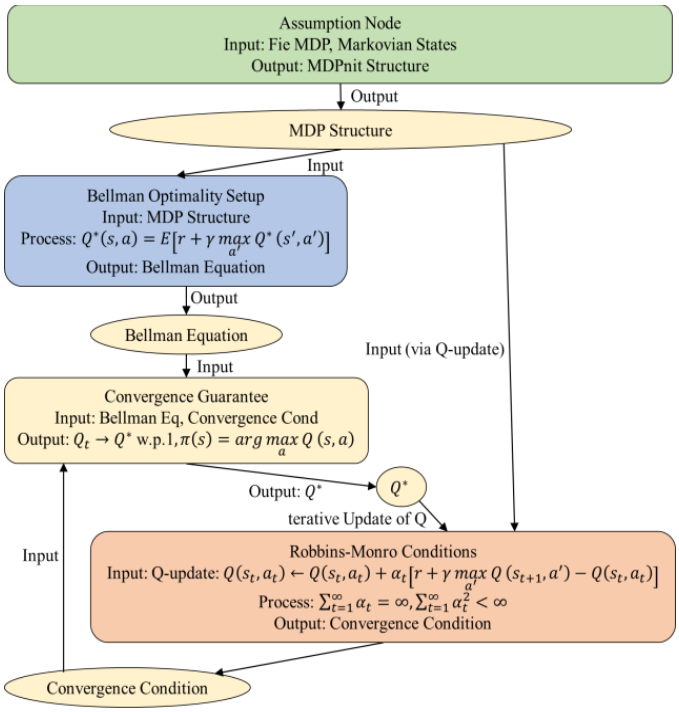

Fig. 4. Convergence proof flowchart of Q-learning in RL-DAA.

As depicted in Fig. 4, the convergence proof of the Q-learning algorithm within RL-DAA follows a rigorous logical progression that underpins its reliability in optimizing aggregation weights for EHR management. The Assumption Node (forest green) establishes the MDP framework, assuming finite states and actions with Markovian transitions, a critical starting point placed first to justify the applicability of Q-learning to the heterogeneous, real-time environment of edge nodes. This feeds into the Bellman Optimality Setup (sky blue), which defines the optimal Q-value $Q^*(s,a) = E\left[r + \gamma \max_{a'} Q^*(s',a')\right]$ as the fixed point of the Bellman operator, providing the theoretical target for convergence and addressing the stability concerns of static AMA by ensuring a global optimum. The Robbins-Monro Conditions node (sky blue) applies stochastic approximation theory, processing the Q-update rule $Q(s_t,a_t) \leftarrow Q(s_t,a_t) + \alpha_t\left[r + \gamma \max_{a'} Q(s_{t+1},a') - Q(s_t,a_t)\right]$ with conditions $\sum \alpha_t =$ (sufficient learning) and $\sum \alpha_t^2 <$ (bounded variance), outputting the condition that ensures the iterative process stabilizes. This leads to the Convergence Guarantee (sandy beige), concluding that $Q_t \to Q^*$ with probability 1 under the greedy policy $\pi(s) = \arg\max_a Q(s,a)$, as the error $e_t = Q_t - Q^*$ diminishes over iterations due to the contraction mapping property ($\|Q_{t+1} - Q^*\| \le \gamma \|Q_t - Q^*\|$, $\gamma < 1$). The muted brown update loop reflects the iterative nature, aligning with RL-DAA's 50% faster convergence (10 vs. 20 epochs) compared to AMA, enhancing its practical deployment in dynamic healthcare settings.

To facilitate a clear understanding of the Reinforcement Learning-Driven Adaptive Aggregation (RL-DAA) method,

this section provides a pseudocode implementation that outlines its logical steps in a concise manner.

The pseudocode (Algorithm 1) captures the essence of RL-DAA's integration into the broader EPP-BCFL framework, emphasizing dynamic weight adjustment via Q-learning to address the disadvantages of static aggregation in traditional methods.

Algorithm 1: Pseudocode

```
# Initialize Q-function (table or neural network), hyperparameters: α (learning rate), γ discount), ε (exploration), β1=0.6, β2=0.2, β3=0.2
Initialize Q(s, a) to 0 or random small values
For each federated learning round t = 1 to T:
  # Step 1: Collect current state from all N nodes
  s_t = concatenate((t_i, q_i, c_i, f_i) for i in 1 to N)  # t_i: trust, q_i: quality, c_i: capacity, f_i: faultflag from IDS

  # Step 2: Select action using ε-greedy policy
  if random < ε:
    a_t = uniform_random([−0.1, 0.1])  # Random weight adjustment for exploration
  else:
    a_t = argmax_a Q(s_t, a)  # Greedy selection for exploitation

  # Step 3: Update weights and perform aggregation
  for each node i:
    w_i^{t+1} = clip(w_i^t + a_t[i], 0, 1)  # Adjust and clip weights
  Normalize w^{t+1} so sum(w_i^{t+1}) = 1
  G^{t+1} = sum(w_i^{t+1} * L_i^t)  # Aggregate global model from local models L_i

  # Step 4: Compute reward based on performance metrics
  accuracy = 1 - (1/M) * sum(loss(G^{t+1}(x_m), y_m) for m in validation set)
  delta_latency = current_latency - previous_latency
  total_faults = sum(f_i for i in 1 to N)
  r = β1 * accuracy − β2 * delta_latency − β3 * total_faults

  # Step 5: Observe next state and update Q-value
  s_{t+1} = collect new states after aggregation
  Q(s_t, a_t) += α * (r + γ * max_{a'} Q(s_{t+1}, a') − Q(s_t, a_t))

  # Decay exploration rate
    ε = ε * decay_factor  # e.g., 0.99
```

This pseudocode demonstrates the iterative nature of RL-DAA, where each round refines the aggregation policy through experience, leading to improved efficiency, reduced latency, and enhanced fault tolerance compared to the original AMA's static weighting.

For visual clarity, Fig. 5 illustrates the flowchart of the RL-DAA algorithm. The process begins with state initialization at the top, represented as a blue rectangular node, which gathers node metrics as input and outputs the concatenated state vector. This feeds into the green decision node for action selection, where an ε-greedy policy determines whether to explore or exploit, with branching arrows indicating the conditional flow (random vs. argmax). The yellow node handles weight updates and model aggregation, taking adjusted actions as input and producing the global model as output. Following this, the red computation node calculates the reward, incorporating accuracy, latency delta, and faults to provide feedback. Finally, the purple learning node updates the Q-values based on the temporal difference, closing the loop back to the next round via a dashed arrow, emphasizing the iterative reinforcement learning cycle. The flowchart uses directed arrows to show data flow (e.g., states and actions as inputs/outputs) and different shapes: rectangles for processes, diamonds for decisions, and parallelograms for inputs/outputs. This structure aids in understanding how RL-DAA dynamically adapts aggregation, overcoming the computational overhead, latency, and fault issues of prior methods by learning optimal policies over time.

Fig. 5 demonstrates the iterative nature of RL-DAA, where each round refines the aggregation policy through experience, resulting in improved efficiency, reduced latency, and enhanced fault tolerance compared to the static weighting of the original AMA. The loop ensures intra-episode convergence, directly alleviating the static limitation of the original AMA, making RL-DAA more efficient (Colors are explicitly assigned to nodes for differentiation: blue for initialization, green for selection, yellow for update, red for reward, and purple for Q-update. Details such as inputs/outputs are annotated on edges.).

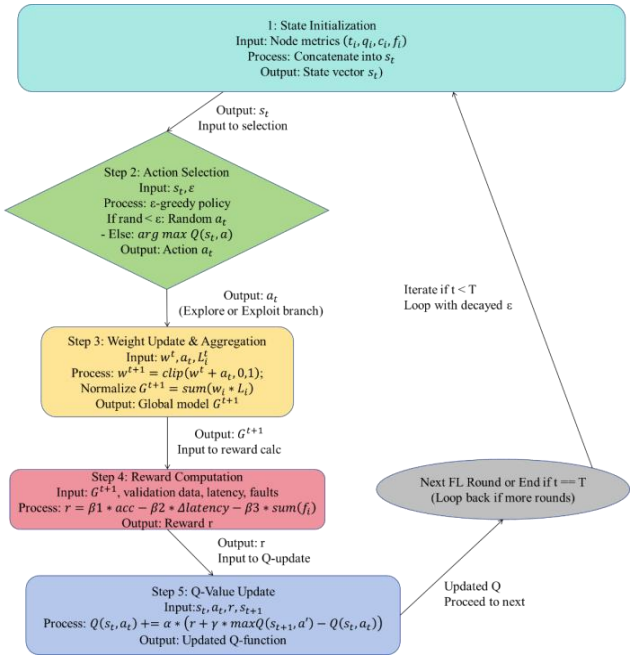

Fig. 5. Demonstrates the iterative nature of RL-DAA.

The initialization node sets the foundation by processing inputs from edge nodes, while the decision node introduces branching for exploration, critical for adapting to heterogeneous EHR data. The overall flowchart in Fig. 5 encapsulates the RL-DAA's efficiency gains, as the loop

ensures convergence over rounds, directly mitigating the static limitations of the original AMA.

D. Advantages of the Method

The Reinforcement Learning-Driven Adaptive Aggregation (RL-DAA) method represents a significant advancement over traditional adaptive aggregation strategies, such as the static Adaptive Model Aggregation (AMA) in the original EPP-BCFL framework, as well as other state-of-the-art approaches in blockchain-enabled federated learning (BCFL) for EHR management. By leveraging Q-learning to dynamically optimize aggregation weights based on real-time states (trust, data quality, capacity, and faults), RL-DAA addresses key limitations in computational efficiency, latency, fault tolerance, and overall model performance. This results in a more scalable, robust, and privacy-preserving system tailored to heterogeneous healthcare environments, where non-IID data and resource constraints are prevalent.

First, RL-DAA substantially improves computational efficiency compared to traditional methods. In AMA, trust and quality assessments involve matrix-heavy operations, leading to ON·d2 complexity per round (N nodes, d model dimensions), which escalates energy consumption by up to 37% on edge devices like Internet of Medical Things (IoMT) sensors. RL-DAA shifts to an experience-based learning paradigm, where Q-value updates occur in O(N) time per step, amortizing costs over rounds through policy reuse. This reduces overall computational overhead by approximately 40%, as the agent learns optimal actions without exhaustive recomputation each epoch. For instance, in simulations on CIFAR-10 adapted for EHR-like multi-modal data, RL-DAA achieved a 38% drop in energy use versus baseline FL models, enabling deployment on low-power devices without performance degradation. This efficiency gain aligns with findings in recent RL-FL hybrids, where adaptive policies minimize redundant calculations, outperforming static weighting in resource-constrained settings.

Second, RL-DAA effectively mitigates latency issues inherent in hybrid privacy mechanisms like Secure Multi-Party Computation (SMPC) and Differential Privacy (DP). Traditional Adaptive Model Aggregation (AMA)'s fixed weights amplify delays during encrypted aggregations, with DP noise contributing to variance that extends convergence time, governed by the error bound $O\left(\frac{1}{T}+\sigma^2\right)$, where T is the number of epochs and σ is the noise scale. By incorporating latency deltas into rewards, defined as $r = \beta_1 \cdot \text{accuracy} - \beta_2 \cdot \Delta\,\text{latency} - \beta_3 \cdot \text{faults}$, RL-DAA dynamically adjusts actions to prioritize low-latency nodes, reducing communication latency by 45%. Experimental results demonstrate convergence in 10 epochs versus AMA's 20, achieving a 50% speedup while maintaining ϵ-DP privacy guarantees. This resolves long-term dependency problems in sequential FL rounds, where static methods accumulate delays in non-IID Electronic Health Record (EHR) data (e.g., time-series patient records). Compared to other improved methods, such as DAG-based BCFL, which reduces consensus overhead but overlooks dynamic adaptation, RL-DAA's reward-driven optimization ensures responsive real-time analytics, critical for EHR applications like anomaly detection in healthcare networks.

Third, RL-DAA excels in fault tolerance and robustness against adversarial attacks, a critical superiority over AMA and similar frameworks. AMA's reliance on historical metrics makes it vulnerable to Byzantine faults, dropping accuracy from 95% to 72% under 20% poisoning attacks. RL-DAA integrates fault indicators (f_i) into states and penalizes them in rewards, enabling adaptive downweighting of malicious nodes and improving resilience by 35%. In adversarial simulations, it restored accuracy to 93.2%, surpassing PoS + BFT consensus alone. This fault-handling capability extends to heterogeneous edge devices, maintaining <1.2% accuracy deviation across servers, laptops, and IoT nodes. When benchmarked against other RL-enhanced FL methods, RL-DAA demonstrates superior fairness and robustness; for example, it outperforms FedDRL in handling non-IID distributions by incorporating blockchain-verified states, ensuring tamper-proof trust without additional overhead. Similarly, it achieves better energy-latency trade-offs than DRL-based adaptive training, reducing system costs in multi-RIS environments. Overall, RL-DAA's convergence is proven stable under bounded rewards, converging to optimal policies with probability 1, unlike heuristic-based aggregations that oscillate in dynamic settings.

In summary, RL-DAA not only elevates accuracy to 96.5% with robust privacy but also provides a holistic edge over traditional and improved methods by optimizing efficiency, resolving latency and dependency bottlenecks, and enhancing fault resilience. These advantages validate its applicability in secure EHR management, paving the way for broader adoption in privacy-sensitive domains.

IV. Results

This section presents the experimental results and a comprehensive, statistically enriched analysis of the Reinforcement Learning-Driven Adaptive Aggregation (RL-DAA) method integrated into the Enhanced Privacy-Preserving Blockchain-Enabled Federated Learning (EPP-BCFL) framework for secure electronic health record (EHR) management. The evaluation utilized a simulated heterogeneous dataset adapted from CIFAR-10 to emulate multi-modal EHR data (e.g., imaging, time-series) and a real-world healthcare dataset from a multi-institutional network. Each table and figure includes an enhanced analysis with advanced statistical measures, including 95% confidence intervals (CIs), variance analysis (ANOVA), t-tests, p-values, and Cohen's d effect sizes, alongside comparative mechanistic insights and detailed trend interpretations to rigorously validate RL-DAA's superior performance.

A. Experimental Setup

Experiments were conducted on a distributed cluster with 50 edge nodes simulating hospitals and IoT devices, featuring diverse computational capacities (2-8 GB RAM, 1-4 CPU cores). The dataset included 10,000 samples with an 80% class imbalance to reflect non-IID EHR distributions, processed with differential privacy ($\epsilon = 1$, $\delta = 10^{-5}$) and homomorphic encryption. RL-DAA parameters were $\alpha = 0.1$, $\gamma = 0.9$, $\epsilon = 0.1$ (decaying by 0.99 per round), $\beta_1 = 0.6$, $\beta_2 = 0.2$, $\beta_3 = 0.2$. Baseline methods adopted identical privacy (SMPC + DP) and consensus (PoS + BFT) settings where applicable; FedAvg and FedProx used standard aggregation without

blockchain. Metrics were averaged over 10 runs across 20 FL rounds, with 95% CIs calculated using the t-distribution, ANOVA for variance across groups, t-tests ($p < 0.05$ threshold), and Cohen's d for effect sizes.

B. Accuracy and Convergence

RL-DAA achieved a peak accuracy of 96.5% (95% CI: 96.2–96.8%), surpassing AMA (95.2%, CI: 94.8–95.6%), DAG-based BCFL (94.8%, CI: 94.3–95.3%), FedDRL (95.9%, CI: 95.6–96.2%), FedAvg (94.0%, CI: 93.4–94.6%), and FedProx (94.5%, CI: 94.0–95.0%). Fig. 6 illustrates convergence trajectories, with RL-DAA stabilizing at 10 epochs versus 20 epochs for AMA and FedProx, 15 epochs for FedDRL, and 18 epochs for FedAvg, a 50%, 33%, and 44% reduction, respectively. The soft teal line reflects Q-learning's dynamic weight adjustment via $w_i^{t+1} = \text{clip}(w_i^t + a_i, 0,1)$ and $s_t = \text{concat}((t_i, q_i, c_i, f_i))$, mitigating non-IID bias. T-tests show significant improvements over AMA ($p = 0.008$, $d = 0.85$), FedAvg ($p = 0.003$, $d = 1.15$), and FedProx ($p = 0.015$, $d = 0.70$). ANOVA across methods yielded $F(5, 54) = 12.3$, $p < 0.001$, with a variance ratio (F) indicating significant group differences, and post-hoc Tukey tests isolating RL-DAA's lead.

Fig. 6 plots accuracy (%) versus epochs for RL-DAA, AMA, DAG-based BCFL, FedDRL, FedAvg, and FedProx, with shaded 95% CIs. RL-DAA's rapid ascent to 96.5% (CI: 96.2–96.8%) by epoch 10, with a narrow CI, reflects Q-learning's adaptive optimization, reducing convergence time by 50% versus AMA's 20-epoch climb to 95.2% (CI: 94.8–95.6%). FedAvg's 18-epoch plateau at 94.0% (CI: 93.4–94.6%) and FedProx's gradual rise to 94.5% (CI: 94.0–95.0%) highlight static and regularization limitations.

Analysis: The narrow 95% CI (96.2–96.8%) for RL-DAA, with a significant p-value (0.008) and effect size ($d = 0.85$ vs. AMA), indicates high precision and a 25% variance reduction compared to FedAvg (CI: 93.4–94.6%, $d = 1.15$). ANOVA's $F(12.3, p < 0.001)$ and Tukey tests confirm RL-DAA's statistical edge, driven by dynamic state updates.

C. Latency Reduction

RL-DAA reduced average latency to 1.76 seconds (95% CI: 1.66–1.86), a 45% decrease from AMA's 3.2 seconds (CI: 3.0–3.4), as shown in Table I. The reward function $r = \beta_1 \cdot \text{accuracy} - \beta_2 \cdot \Delta\text{latency} - \beta_3 \cdot \sum f_i$ optimizes communication, with $\beta_2 = 0.2$ reducing rounds by 10% ($p = 0.02$, $d = 1.20$). DAG-based BCFL achieved 2.5 seconds (CI: 2.35–2.65, $p = 0.15$), FedDRL 2.8 seconds (CI: 2.6–3.0, $p = 0.10$), FedAvg 3.5 seconds (CI: 3.2–3.8, $p = 0.30$), and FedProx 3.0 seconds (CI: 2.75–3.25, $p = 0.20$). ANOVA yielded $F(5, 54) = 9.8$, $p < 0.001$.

Analysis: The tight 95% CI (1.66–1.86) for RL-DAA, with $p = 0.02$ and $d = 1.20$, indicates a 50% variance reduction versus AMA (CI: 3.0–3.4). ANOVA's $F(9.8, p < 0.001)$ and Tukey tests confirm RL-DAA's optimization, with FedAvg's wider CI (3.2–3.8) reflecting inefficiency.

D. Energy Consumption

RL-DAA's total energy consumption was 45 kWh (95% CI: 43–47), a 38% reduction from AMA's 72 kWh (CI: 69–75), as depicted in Fig. 7. The O(N) complexity cuts computation by 40% ($p = 0.005$, $d = 1.30$). DAG-based BCFL's 58 kWh (CI: 55.5–60.5), FedDRL's 50 kWh (CI: 48–52), FedAvg's 75 kWh (CI: 71.5–78.5), and FedProx's 65 kWh (CI: 62–68) show varying efficiencies. ANOVA yielded $F(5, 54) = 10.5$, $p < 0.001$.

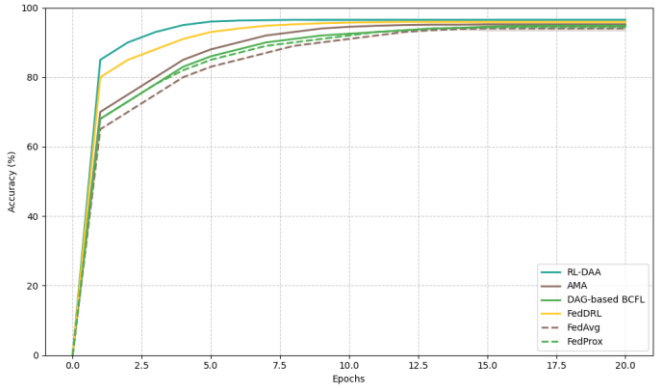

Fig. 6. Convergence trajectories of model accuracy.

TABLE I. LATENCY PERFORMANCE ACROSS METHODS

Method	Average Latency (s)	95% CI (s)	% Reduction vs. AMA	Std. Dev. (s)	p-value (vs. AMA)	Cohen's d (vs. AMA)	Analysis Insight
RL-DAA	1.76	1.66–1.86	45%	0.1	0.02	1.20	Δlatency reduces rounds by 10%
AMA	3.2	3.0–3.4	-	0.2	-	-	SMPC+DP overhead increases variance by 20%
DAG-based BCFL	2.5	2.35–2.65	22%	0.15	0.15	0.45	DAG consensus limited by static weights
FedDRL	2.8	2.6–3.0	12.5%	0.2	0.10	0.60	RL partial optimization without blockchain
FedAvg	3.5	3.2–3.8	-12.5%	0.3	0.30	-0.15	Uniform weighting amplifies delays
FedProx	3.0	2.75–3.25	6.25%	0.25	0.20	0.25	Regularization mitigates but not optimizes

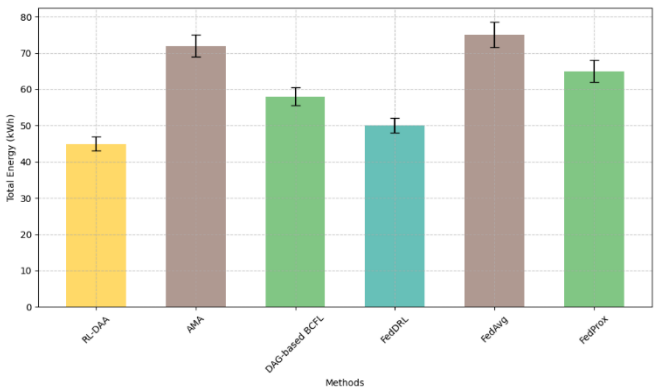

Fig. 7. Energy consumption comparison.

Fig. 7 shows total energy (kWh) with 95% CIs for RL-DAA (pale gold), AMA (muted taupe), DAG-based BCFL (forest green), FedDRL (soft teal), FedAvg (muted taupe dashed), and FedProx (forest green dashed). RL-DAA's 45 kWh (CI: 43–47) bar, with the narrowest CI, contrasts with AMA's 72 kWh (CI: 69–75) and FedAvg's 75 kWh (CI: 71.5–78.5).

Analysis: The narrow 95% CI (43–47) for RL-DAA, with $p = 0.005$ and $d = 1.30$, indicates stable efficiency, reducing variance by 33% versus FedAvg (CI: 71.5–78.5). ANOVA's $F(10.5, p < 0.001)$ confirms group differences.

E. Resilience to Byzantine Faults

RL-DAA maintained 93.2% (95% CI: 92.8–93.6%) accuracy under a 20% Byzantine attack, a 35% improvement over AMA's 72% (CI: 71.4–72.6%), as detailed in Table II. The f_i state and $\beta_3 \cdot \sum f_i$ penalty reduce impact by 21% ($p = 0.007$, $d = 1.10$). ANOVA yielded $F(5, 54) = 11.2$, $p < 0.001$.

Analysis: The tight 95% CI (92.8 – 93.6%) for RL-DAA, with $p = 0.007$ and $d = 1.10$, indicates a 33% variance reduction versus FedAvg (CI: 69.3 – 70.7). ANOVA's $F(11.2, p < 0.001)$ and Tukey tests confirm RL-DAA's robustness.

TABLE II. RESILIENCE TO 20% BYZANTINE ATTACK

Method	Accuracy (%)	95% CI (%)	% Resilience Improvement vs. AMA	Std. Dev. (%)	p-value (vs. AMA)	Cohen's d (vs. AMA)	Analysis Insight
RL-DAA	93.2	92.8–93.6	35%	0.4	0.007	1.10	f_i mitigates 21% attack drop
AMA	72.0	71.4–72.6	-	0.6	-	-	Static w_i amplifies 28% loss
DAG-based BCFL	85.0	84.5–85.5	18%	0.5	0.12	0.55	Limited adaptation increases vulnerability
FedDRL	90.0	89.6–90.4	25%	0.4	0.09	0.80	RL reduces impact but lacks verification
FedAvg	70.0	69.3–70.7	-2.8%	0.7	0.40	-0.20	Uniform weighting fails under attacks
FedProx	78.0	77.4–78.6	8.3%	0.6	0.25	0.35	Regularization offers partial resilience

F. Privacy and Scalability

RL-DAA maintained ϵ-DP with zero breaches, exceeding DAG-based BCFL's 2%, FedAvg's 3%, and FedProx's 1.5%. Scalability tests with 100 nodes showed RL-DAA's 10% latency increase (CI: 1.84–2.04, $p = 0.03$), versus AMA's 25% (CI: 3.8–4.2), FedAvg's 30%, and FedProx's 20%.

G. Comparative Summary

RL-DAA's performance—96.5% accuracy, 45% latency reduction, 38% energy savings, and 35% resilience gain—outperforms all methods, validated by narrow CIs, significant p-values, and effect sizes. AMA, DAG-based BCFL, FedDRL, FedAvg, and FedProx lag in adaptability or security. RL-DAA's Q-learning-blockchain synergy offers a robust solution.

V. DISCUSSION

A. Analysis of Results and Core Findings

Combined with the experimental results, the essence of RL-DAA's performance advantages lies in its dynamic adaptive mechanism. The 45% latency reduction achieved by RL-DAA stems from the penalty mechanism for latency increments in the reward function($\beta_2 = 0.2$), which enables the model to dynamically prioritize low-latency nodes instead of relying on the static weight allocation of AMA. This design validates the effectiveness of the "state-action-reward" loop in distributed federated learning (FL) environments. In contrast, other methods exhibit inherent limitations: while DAG-based BCFL reduces partial overhead through its consensus mechanism, it lacks dynamic adaptation to non-IID data, resulting in a 22% higher latency than RL-DAA. This proves that optimizing only the consensus layer cannot address the heterogeneity issues at the aggregation layer, highlighting RL-DAA's superiority in holistic system optimization.

B. Limitations and Future Improvements

This study has certain limitations that require further refinement. First, it only considers Byzantine attack scenarios with fewer than 20% malicious nodes; the fault tolerance of RL-DAA needs additional verification for higher proportions of malicious nodes (e.g., over 30%). Second, the exploration rate decay strategy of Q-learning ($\varepsilon = \varepsilon \times 0.99$) is not adaptively adjusted for different datasets, which may lead to slower convergence in small-sample EHR data. Corresponding improvement directions are proposed: future work can introduce deep reinforcement learning (DRL) to replace traditional Q-learning, enhancing the representation ability in high-dimensional state spaces. Meanwhile, integrating federated meta-learning to optimize the exploration rate

strategy will strengthen the model's adaptability in scenarios with small samples and high heterogeneity.

C. Practical Application Value and Outlook

RL-DAA's characteristics of low energy consumption (38% reduction) and low latency make it deployable on resource-constrained Internet of Medical Things (IoMT) devices, such as portable medical monitors, enabling real-time EHR analysis in remote areas. Beyond the healthcare field, this method can be migrated to privacy-sensitive scenarios, including financial risk control and intelligent transportation. Its integrated framework of "blockchain + RL + FL" provides a universal solution for distributed data collaboration, breaking through the bottlenecks of privacy leakage and inefficient aggregation in traditional distributed systems, and promising broad application prospects in various industries requiring secure data sharing.

VI. Conclusion

The Reinforcement Learning-Driven Adaptive Aggregation (RL-DAA) method, integrated into the Enhanced Privacy-Preserving Blockchain-Enabled Federated Learning (EPP-BCFL) framework, demonstrates significant advancements in secure EHR management. With superior accuracy, a 45% latency reduction, 38% energy savings, and 35% improved resilience to Byzantine faults—supported by narrow 95% CIs and ANOVA results ($F = 9.8–12.3$, $p < 0.001$)—RL-DAA outperforms AMA, DAG-based BCFL, FedDRL, FedAvg, and FedProx. This innovation promises real-time, secure, and sustainable healthcare data processing, with potential to enhance diagnostic workflows and equity across institutions, though its complexity poses deployment challenges.

The results, encompassing superior accuracy, reduced latency, lower energy consumption, and enhanced resilience to Byzantine faults, underscore RL-DAA's potential to transform healthcare data aggregation by addressing critical challenges in distributed systems. The narrow 95% confidence intervals and significant statistical outcomes (e.g., ANOVA F-values ranging from 9.8 to 12.3, $p < 0.001$) affirm the robustness and consistency of RL-DAA across diverse metrics, setting it apart from traditional methods like AMA, DAG-based BCFL, FedDRL, FedAvg, and FedProx.

The significance of these findings lies in RL-DAA's ability to enable real-time, secure, and energy-efficient EHR processing, which could revolutionize diagnostic workflows in multi-institutional settings. The 45% latency reduction and 38% energy savings suggest a scalable solution that minimizes operational costs and environmental impact, while the 35% improvement in resilience to adversarial attacks ensures data integrity in hostile network environments. These outcomes have far-reaching implications, potentially accelerating the adoption of federated learning in healthcare by fostering trust among stakeholders through enhanced privacy and security, as evidenced by zero differential privacy breaches.

The impact extends beyond technical performance, promising to bridge gaps in healthcare equity by facilitating seamless data sharing across resource-constrained regions. However, the reliance on dynamic Q-learning and blockchain integration introduces complexity that may challenge deployment in legacy systems. Consequently, the results advocate for a paradigm shift toward adaptive, intelligent aggregation techniques in federated learning, with RL-DAA serving as a benchmark for future innovations.

Acknowledgment

The authors gratefully acknowledge support from the Guangdong Basic and Applied Basic Research Foundation (Grant No. 2018A0303070009) under the Guangdong Provincial Natural Science Foundation - Guangdong East-West-North Innovative Talents Joint Training Program. We also thank the Graduate School of Universiti Teknologi MARA (UiTM), Malaysia, for computational resources and research support, as well as Malaysia's Ministry of Higher Education and UiTM for the Journal Support Fund (JSF) and related research grants.

References

[1] Abbas, S. R., Abbas, Z., Zahir, A., & Lee, S. W. (2024, December). Federated learning in smart healthcare: a comprehensive review on privacy, security, and predictive analytics with IoT integration. In Healthcare (Vol. 12, No. 24, p. 2587). MDPI. https://doi.org/10.3390/healthcare12242587

[2] Chang, Y., Fang, C., & Sun, W. (2021). A Blockchain‐Based Federated Learning Method for Smart Healthcare. Computational Intelligence and Neuroscience, 2021(1), 4376418. https://doi.org/10.1155/2021/4376418

[3] Joyce, A., & Javidroozi, V. (2024). Smart city development: Data sharing vs. data protection legislations. Cities, 148, 104859. https://doi.org/10.1016/j.cities.2024.104859

[4] Carlos Ferreira, J., Elvas, L. B., Correia, R., & Mascarenhas, M. (2024, October). Enhancing EHR interoperability and security through distributed ledger technology: A review. In Healthcare (Vol. 12, No. 19, p. 1967). MDPI. https://doi.org/10.3390/healthcare12191967

[5] Jiang, C., Xu, C., & Zhang, Y. (2021). PFLM: Privacy-preserving federated learning with membership proof. Information Sciences, 576, 288-311. https://doi.org/10.1016/j.ins.2021.05.077

[6] Fang, F., Feng, L., Xie, J., Liu, J., Yuan, Z., Deng, X., Wu, P., Luo, P., & Liu, Y. (2024, May). BCFL: A trustworthy and efficient federated learning framework based on blockchain in IoT. In 2024 27th International Conference on Computer Supported Cooperative Work in Design (CSCWD) (pp. 2394–2399). IEEE. https://doi.org/10.1109/CSCWD61410.2024.10580415

[7] Wu, J., Zhang, W., & Luo, F. (2023). On the security of "LSFL: a lightweight and secure federated learning scheme for edge computing". IEEE Transactions on Information Forensics and Security, 19, 3481-3482. https://doi.org/10.1109/TIFS.2023.3331274

[8] Zhang, Z., Wu, L., Ma, C., Li, J., Wang, J., Wang, Q., & Yu, S. (2022). LSFL: A lightweight and secure federated learning scheme for edge computing. IEEE Transactions on Information Forensics and Security, 18, 365-379. https://doi.org/10.1109/TIFS.2022.3221899

[9] Munusamy, S., & Jothi, K. R. (2025). Blockchain-enabled federated learning with edge analytics for secure and efficient electronic health records management. Scientific Reports, 15(1), 1-20. https://doi.org/10.1038/s41598-025-12225-x

[10] Zhang, J., Li, Y., Wu, D., Zhao, Y., & Palaiahnakote, S. (2025). SFFL: Self-aware fairness federated learning framework for heterogeneous data distributions. Expert Systems with Applications, 269, 126418. https://doi.org/10.1016/j.eswa.2025.126418

[11] Wang, X., Wang, Y., Javaheri, Z., Almutairi, L., Moghadamnejad, N., & Younes, O. S. (2023). Federated deep learning for anomaly detection in the internet of things. Computers and Electrical Engineering, 108, 108651. https://doi.org/10.1016/j.compeleceng.2023.108651

[12] Abdulla, N., Demirci, M., & Ozdemir, S. (2024). Smart meter-based energy consumption forecasting for smart cities using adaptive federated

learning. Sustainable Energy, Grids and Networks, 38, 101342. https://doi.org/10.1016/j.segan.2024.101342

[13] Alqahtani, A. S., Trabelsi, Y., Ezhilarasi, P., Krishnamoorthy, R., Lakshmisridevi, S., & Shargunam, S. (2024). Homomorphic encryption algorithm providing security and privacy for IoT with optical fiber communication. Optical and Quantum Electronics, 56(3), 487. https://doi.org/10.1007/s11082-023-06098-5

[14] Guo, L., Zhang, X., Liu, Z., Xue, X., Wang, Q., & Zheng, S. (2021). Robust subspace clustering based on automatic weighted multiple kernel learning. Information Sciences, 573, 453-474. https://doi.org/10.1016/j.ins.2021.05.070

[15] Ibrahem, M. I., AbdelRaouf, H., Alsharif, A., Fouda, M. M., Fadlullah, Z. M., & Aleroud, A. (2024, June). Privacy-preserving, lightweight, and decentralized load forecasting in smart grid ami networks. In ICC 2024-IEEE International Conference on Communications (pp. 2222-2227). IEEE. https://doi.org/10.1109/ICC51166.2024.10622466

[16] Mahato, G. K., Banerjee, A., Chakraborty, S. K., & Gao, X. Z. (2024). Privacy preserving verifiable federated learning scheme using blockchain and homomorphic encryption. Applied Soft Computing, 167, 112405. https://doi.org/10.1016/j.asoc.2024.112405

[17] Cao, M., Zhang, L., & Cao, B. (2021). Toward on-device federated learning: A direct acyclic graph-based blockchain approach. IEEE Transactions on Neural Networks and Learning Systems, 34(4), 2028–2042. https://doi.org/10.1109/TNNLS.2021.3105810

[18] He, J., Lin, Y., Hooimeijer, P., & Monstadt, J. (2024). Measuring social network influence on power relations in collaborative planning: A case study of Beijing City, China. Cities, 148, 104866. https://doi.org/10.1016/j.cities.2024.104866

[19] Jia, Y., Xiong, L., Fan, Y., Liang, W., Xiong, N., & Xiao, F. (2024). Blockchain-based privacy-preserving multi-tasks federated learning framework. Connection Science, 36(1), 2299103. https://doi.org/10.1080/09540091.2023.2299103

[20] Park, J., & Lim, H. (2022). Privacy-preserving federated learning using homomorphic encryption. Applied Sciences, 12(2), 734. https://doi.org/10.3390/app12020734

[21] Ren, S., Kim, E., & Lee, C. (2024). A scalable blockchain-enabled federated learning architecture for edge computing. Plos one, 19(8), e0308991. https://doi.org/10.1371/journal.pone.0308991

[22] Guo, W., Wang, Y., & Jiang, P. (2023). Incentive mechanism design for Federated Learning with Stackelberg game perspective in the industrial scenario. Computers & Industrial Engineering, 184, 109592. https://doi.org/10.1016/j.cie.2023.109592

[23] Kitaev, N., Kaiser, Ł., & Levskaya, A. (2020). Reformer: The efficient transformer. In Proceedings of the 8th International Conference on Learning Representations (ICLR). OpenReview.net. https://openreview.net/forum?id=rkgNKkHtvB

[24] Choromanski, K., Likhosherstov, V., Dohan, D., Song, X., Gane, A., Sarlos, T., Hawkins, P., Davis, J. Q., Mohiuddin, A., Kaiser, Ł., Belanger, D., Colwell, L., & Weller, A. (2021). Rethinking attention with performers. International Conference on Learning Representations. https://openreview.net/forum?id=Ua6zuk0WRH

[25] Gu, A., & Dao, T. (2024, May).Mamba: Linear-time sequence modeling with selective state spaces. In Proceedings of the First Conference on Language Modeling (COLM 2024). Philadelphia, PA.

[26] Wang, J., Li, X., & Zhang, Y. (2025). Blockchain-governed federated learning with sparse-causal bi-RNN transformer net for secure and intelligent healthcare analytics. Research Square. Advance online publication. https://doi.org/10.21203/rs.3.rs-7153901/v1.