



Research article

Privacy-preserving gradient scheme for hot-rolled steel based on functional encryption

Yadong Wang^{1,2}, Nuerken Saireke^{1,2} and Jimin Wang^{1,2,*}

¹ School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing 100083, China

² Key Laboratory of Knowledge Automation for Industrial Processes, Ministry of Education, Beijing 100083, China

* **Correspondence:** Email: jimwang@ustb.edu.cn.

Abstract: To address the vulnerability of output-layer gradient leakage during the distributed training of hot-rolled steel strip-crown prediction across multiple steel mills, this paper proposes a privacy-preserving split-layer federated learning scheme based on multi-client functional encryption. The proposed scheme keeps the input layer parameters at local clients, where they are updated using local data but are neither uploaded nor aggregated. Only the output layer gradients are encrypted and aggregated to update the shared output layer parameters. Leveraging the inner-product property of multiclient functional encryption alongside a capacity-aware, fixed client partitioning strategy, the secure aggregation of output layer gradients is achieved entirely within the ciphertext domain. Theoretical analysis derives a conditional stationarity bound for the shared output layer optimization phase under fixed local input layer parameters and quantization error. Furthermore, a reduction-based security analysis establishes single-round and multi-round ciphertext indistinguishability under the condition that all authorized fixed-subset outputs are identical between the compared inputs in every communication round. Under the honest-but-curious aggregation server and external eavesdropper threat models, this result provides confidentiality beyond the information revealed by these authorized outputs. Experimental results demonstrate that the proposed secure aggregation framework based on multi-client functional encryption achieves a root mean square error of 3.6290 μm for strip-crown prediction. This result confirms the feasibility of applying privacy-preserving gradient aggregation to this industrial prediction task.

Keywords: hot-rolled steel; functional encryption; privacy preservation; federated learning

1. Introduction

Hot-rolled steel is a foundational material in modern industries, underpinning critical applications in construction, automotive manufacturing, and national key projects [1, 2]. Its mechanical properties—such as strength, toughness, and flatness (e.g., crown)—directly determine product safety, service life, and manufacturing yield. However, traditional hot-rolled steel process development relies heavily on empirical trials, incurring high economic costs and long lead times [3]. Data-driven prediction of hot-rolled steel properties has become a core enabler for optimizing rolling processes, reducing defect rates, and realizing real-time quality control [4, 5]. Early efforts to predict hot-rolled steel properties focused on physical metallurgy models, which established quantitative process-microstructure property relationships [6, 7]. Since the 1990s, machine learning has emerged as a more flexible alternative: Hodgson et al. [8] established a representative work by utilizing artificial neural networks to model the intricate nonlinear relationships between chemical composition, processing parameters, and the final mechanical properties of hot-rolled carbon steels; Mo et al. [9] optimized neural network hyperparameters via particle swarm optimization and genetic algorithms; Li et al. [10] characterized the static softening of Nb-microalloyed steels using an integrated machine learning approach to decouple complex microstructural interactions. These methods have enhanced prediction precision, but they do not address challenges in industrial scenarios, such as data and model privacy. Steel industry data contains proprietary composition systems, process parameters for high-grade alloys, and industrial knowledge graphs [11, 12]. The deep integration of industrial intelligent systems has exposed them to severe cyber-physical threats. To date, security research has predominantly targeted active deception and data injection attacks within networked environments, such as sampled-data control [13], spatial localization [14], and unmanned aerial vehicles [15]. Yet, this heavy focus on active disruptions obscures an equally critical vulnerability: the passive leakage of sensitive industrial data during distributed collaborative modeling. To break data silos while mitigating such passive leakage, federated learning (FL) is a potential solution. In this split-layer FL architecture, clients keep the input layer local. During federated training, each client updates its input layer parameters using local data. These parameters are neither uploaded to the aggregator nor directly involved in federated aggregation, but they affect the shared output layer parameters through the latent features and output layer gradients. Only the output layer gradients are encrypted, uploaded, and aggregated. These gradients may reveal training statistics or the evolution of the shared output layer. In extreme cases, the adversary could even replicate the collaborative decision logic [16]. The convergence analysis considers the shared output layer optimization problem with fixed input layer parameters and does not cover the coupled updates of the input and output layer parameters.

Accordingly, the core privacy vulnerability lies in client-side output-layer gradients, exposed as raw values and exploitable by an honest-but-curious server or external eavesdropper. Though homomorphic encryption (HE), differential privacy (DP), and secure multi-party computation can in principle protect gradients, they suffer practical limits for hot-rolled steel production: HE brings large overhead in industrial real-time scenarios [17]; DP adds privacy-protecting noise that degrades prediction accuracy [18]; multi-party computation demands frequent cross-site interactions, raising latency and failure risks in industrial environments [19].

To overcome these limitations, we adopt a cryptographic primitive suitable for this setting:

multi-client functional encryption (MCFE). It enables the aggregator to directly compute the authorized fixed-subset sums from the encrypted output layer gradients of multiple clients, and the ciphertext-indistinguishability guarantee prevents the ciphertexts from disclosing additional information not determined by these sums. To further handle heterogeneous workshop environments, we introduce a capacity-aware, fixed client partitioning strategy (FCPS). The clients are partitioned once according to their computational capabilities before collaborative training, and the resulting partition remains unchanged throughout all communication rounds. This paper's contributions are threefold:

- 1) Designs a privacy-preserving scheme for hot-rolled steel artificial neural network (ANN) models using MCFE, enabling secure aggregation of output layer gradients over authorized fixed subsets while preserving local input layers. For encoded-gradient collections with the same authorized fixed-subset outputs, the resulting ciphertext collections remain computationally indistinguishable. The input layer parameters participate in local training but are neither uploaded nor aggregated. Only the output layer gradients are aggregated in the ciphertext domain. A conditional stationarity bound is established for the shared output layer optimization problem with fixed input layer parameters under quantization error.
- 2) Designs FCPS to accommodate heterogeneous client capabilities while maintaining controlled subset sizes. The partition is generated once before collaborative training and reused throughout all communication rounds, avoiding repeated partitioning and functional-key updates.
- 3) Validates the scheme on a strip-crown prediction dataset from a steel plant, demonstrating the compatibility between privacy preservation and prediction accuracy. The evaluation reports prediction performance, quantization reconstruction error, and runtime distribution to assess the feasibility of the proposed scheme.

The remainder of this paper is organized as follows. Section 2 introduces the problem formulation. Section 3 presents the MCFE-based secure gradient aggregation algorithm along with the convergence and security analyses. Section 4 reports the simulation results. Finally, Section 5 concludes the paper.

2. Problem formulation

2.1. Hot-rolled steel ANN prediction model and split-layer architecture

To accurately predict the strip crown of hot-rolled steel, this paper constructs a three-layer ANN with a 34–12–1 topology, as shown in Table 1. Specifically, for the k th local sample $u_{i,k} \in \mathbb{R}^{34}$ of client i , the input layer parameterized by ϕ_i^t maps the input to a latent feature $h_{i,k}^t = f(u_{i,k}; \phi_i^t) \in \mathbb{R}^d$, where $d = 12$ denotes the dimension of the shared output layer weight vector. At communication round t , the input layer parameters ϕ_i^t are updated using the local dataset and remain at client i . They are neither uploaded to the aggregator nor directly aggregated. Let $w \in \mathbb{R}^d$ denote the shared output layer weight vector, and let w_t denote its value at communication round t . Given w_t broadcast by the server, the shared output layer maps the latent feature to the final strip-crown prediction $\hat{\tau}_{i,k}^t = w_t^\top h_{i,k}^t$. During the model optimization phase, client i evaluates the local empirical loss function based on its private dataset $\mathcal{D}_i$ of size $|\mathcal{D}_i|$:

$$F_i(w \mid \phi_i^t) = \frac{1}{|\mathcal{D}_i|} \sum_{k=1}^{|\mathcal{D}_i|} (\tau_{i,k} - \hat{\tau}_{i,k}^t)^2, \quad (2.1)$$

where $\tau_{i,k}$ represents the actual strip-crown measurement. To collaboratively optimize the output layer while keeping the input layer parameters ϕ_i^t private, we define the global conditional objective function as the average empirical loss across all N clients:

$$F(w \mid \{\phi_i^t\}) = \frac{1}{N} \sum_{i=1}^N F_i(w \mid \phi_i^t). \quad (2.2)$$

To minimize this objective, each client computes the local gradient with respect to the shared output layer weight vector:

$$g_{i,t} = \nabla_w F_i(w \mid \phi_i^t) \big|_{w=w_t} = \frac{2}{|\mathcal{D}_i|} \sum_{k=1}^{|\mathcal{D}_i|} (w_t^\top h_{i,k}^t - \tau_{i,k}) h_{i,k}^t. \quad (2.3)$$

The ideal global average gradient is mathematically defined as the arithmetic mean of all local gradients,

$$g_t = \frac{1}{N} \sum_{i=1}^N g_{i,t}. \quad (2.4)$$

After securely aggregating the gradients, the server executes the output layer weight vector update:

$$w_{t+1} = w_t - \eta_t g_t, \quad (2.5)$$

where $\eta_t > 0$ denotes the learning rate at communication round t , which controls the step size of the output layer weight vector update.

Table 1. Configuration of the ANN prediction model for hot-rolled steel.

Layer	Neurons	Activation	Physical correspondence / description
Input	34	–	Process parameters, including entry thickness, rolling force, bending force, roll shift, speed, temperature, and other rolling variables.
Hidden	12	Tanh	Feature extraction for nonlinear metallurgical relationships.
Output	1	Linear	Strip-crown prediction value (μm).

2.2. Privacy challenges

In the proposed FL framework, the primary privacy vulnerability stems from the transmission of local gradients $g_{i,t} \in \mathbb{R}^d$. To formally characterize the security perimeter, we consider the following two representative adversary models:

Honest-but-curious aggregation server [20]: The aggregation server strictly adheres to the prescribed computational logic of the protocol without injecting malicious data or deviating from the execution flow. However, it systematically attempts to infer or reconstruct sensitive client-side information by analyzing all intermediate artifacts exposed during the protocol execution, such as gradient ciphertexts or encoded subset sums.

External eavesdropper [20]: This is a passive monitoring entity lurking within the communication channels between clients and the server. By intercepting client-to-aggregation ciphertext transmissions,

public parameters, and round labels, the eavesdropper attempts to infer sensitive information about client-side gradients.

Under an honest-but-curious aggregation server and external eavesdropper threat model, an aggregator or an external adversary could exploit the statistical correlations between gradient patterns and local data distributions [11].

2.3. Multi-client functional encryption

MCFE is a multi-client extension of single-input functional encryption, where encryption operations are performed by multiple independent clients [21]. Following the labeled MCFE formulation in [22], we introduce a public label into each encryption instance to bind ciphertexts to the same communication round. In this paper, the label is chosen as the communication round index, that is,

$$\ell_t = t. \quad (2.6)$$

Specifically, the plaintext data are divided into N independent parts, held respectively by N clients. Let the plaintext data held by the i th client at communication round t be $\mathbf{x}_{i,t} \in \mathbb{Z}_P^d$, where P denotes the plaintext modulus. The ciphertext collection used for aggregation in round t is composed of the ciphertexts from all clients: $\{ct_{i,t}^{\ell_t}\}_{i \in [N]} = \{ct_{1,t}^{\ell_t}, ct_{2,t}^{\ell_t}, \dots, ct_{N,t}^{\ell_t}\}$, where each ciphertext is independently generated by the corresponding client and is bound to the round label ℓ_t . An MCFE scheme consists of the following four algorithms [23]:

- 1) **Setup**($1^\kappa, 1^d, 1^N$) $\rightarrow$ (mpk, msk, $\{sk_i\}_{i \in [N]}$): Taking the security parameter 1^κ , data dimension 1^d , and the total number of clients 1^N as inputs, it outputs the master public key mpk, the master secret key msk, and the client encryption key sk_i ;
- 2) **Encrypt**($sk_i, i, \ell_t, \mathbf{x}_{i,t}$) $\rightarrow ct_{i,t}^{\ell_t}$: Taking the client encryption key sk_i , the client identifier $i \in [N]$, the round label ℓ_t , and plaintext data $\mathbf{x}_{i,t}$ as inputs, it outputs the labeled ciphertext $ct_{i,t}^{\ell_t}$;
- 3) **KeyGen**(msk, $\mathbf{y}$) $\rightarrow dk_y$: Taking the master secret key msk and the function vector $\mathbf{y} \in \mathbb{Z}_P^N$ as inputs, it outputs the functional key dk_y ;
- 4) **Decrypt**($dk_y, \ell_t, \{ct_{i,t}^{\ell_t}\}_{i \in [N]}$) $\rightarrow \mathbf{z}$: Taking the functional key dk_y , the round label ℓ_t , and the ciphertexts with the same label $\{ct_{i,t}^{\ell_t}\}_{i \in [N]}$ as inputs, it outputs the functional aggregation result

$$\mathbf{z} = \sum_{i=1}^N y_i \mathbf{x}_{i,t} \in \mathbb{Z}_P^d. \quad (2.7)$$

If the input ciphertexts are associated with inconsistent labels, the decryption algorithm returns $\perp$.

All algorithms take mpk as an implicit input. In summary, this section introduces MCFE as the cryptographic foundation for computing authorized functional aggregation results from encrypted gradients in ANN-based federated strip-crown prediction. The key privacy risk lies in the transmission of client-side output layer gradients, which may expose sensitive client-side information through statistical correlations between gradient patterns and local data distributions.

3. Main result

3.1. Privacy-preserving gradient aggregation algorithm

To meet the integer domain requirements of the MCFE protocol, this section designs a secure FL framework consisting of a linear encoding–decoding mechanism and a ciphertext aggregation protocol.

3.1.1. Encoding and decoding mechanism

To bridge the continuous floating-point domain of neural network gradients and the discrete integer ring $\mathbb{Z}_P$ of MCFE, we design a linear encoding–decoding mechanism.

Encoder: The j th component $g_{i,t,j}$ of the local gradient vector $g_{i,t}$ is mapped to a strictly non-negative integer $x_{i,t,j} \in \mathbb{Z}_P$, which serves as the encoded gradient component via scaling, rounding, truncation, and shifting:

$$q_{i,t,j} = \text{clip}(-Q_{\max}, \lfloor g_{i,t,j} \times S \rfloor, Q_{\max}), \quad (3.1)$$

$$x_{i,t,j} = q_{i,t,j} + K, \quad (3.2)$$

where $q_{i,t,j}$ denotes the intermediate bounded quantized value of the j th component of $g_{i,t}$. The parameter $S = 2^{13}$ is the scaling factor preserving fractional precision, and $\lfloor \cdot \rfloor$ denotes nearest-integer rounding. The function $\text{clip}(a, v, b) = \max(a, \min(v, b))$ strictly restricts the value within $[-Q_{\max}, Q_{\max}]$, where $Q_{\max} = 2^{15} - 1$. Finally, a positive offset $K = 2^{15}$ is added to safely eliminate any negative values. The resulting plaintext vector constructed from all d components is denoted as $\mathbf{x}_{i,t} = (x_{i,t,1}, x_{i,t,2}, \dots, x_{i,t,d})^\top$, which serves as the concrete input to the labeled MCFE encryption algorithm under the round label ℓ_t in round t .

Decoder: Let U_ℓ denote the ℓ th fixed authorized aggregation subset defined before collaborative training. For each subset U_ℓ , define the encoded subset-sum vector as

$$X_{U_\ell,t} = \sum_{i \in U_\ell} \mathbf{x}_{i,t} = (X_{U_\ell,t,1}, X_{U_\ell,t,2}, \dots, X_{U_\ell,t,d})^\top, \quad (3.3)$$

where

$$X_{U_\ell,t,j} = \sum_{i \in U_\ell} x_{i,t,j}, \quad j = 1, 2, \dots, d. \quad (3.4)$$

The decoder then performs the inverse mapping to recover the quantized subset gradient sum

$$G_{U_\ell,t,j} = \frac{X_{U_\ell,t,j} - |U_\ell| \cdot K}{S}, \quad (3.5)$$

where $|U_\ell|$ denotes the number of clients in the authorized fixed subset. The recovered subset gradient sum vector is therefore $G_{U_\ell,t} = (G_{U_\ell,t,1}, G_{U_\ell,t,2}, \dots, G_{U_\ell,t,d})^\top$.

3.1.2. Secure aggregation protocol

Building upon the encoding–decoding mechanism, this section presents the secure aggregation protocol. To handle heterogeneous workshop environments while keeping client membership unchanged across communication rounds, our framework employs FCPS. Before collaborative training, a capacity score is calculated once for each client $i = 1, 2, \dots, N$ as

$$\rho_i = \alpha \widehat{C}_i + \beta \widehat{M}_i - \gamma \widehat{R}_i. \quad (3.6)$$

Here, $\widehat{C}_i$, $\widehat{M}_i$, and $\widehat{R}_i$ denote the normalized CPU availability, available memory, and historical response latency measured before collaborative training, respectively. The weights α , β , and γ are non-negative and satisfy $\alpha + \beta + \gamma = 1$.

FCPS sorts all clients once according to ρ_i and partitions adjacent clients into N_{sub} nonempty and mutually disjoint fixed subsets. The resulting fixed partition is denoted by

$$\Pi = \{U_1, U_2, \dots, U_{N_{\text{sub}}}\}, \quad (3.7)$$

where N_{sub} denotes the number of fixed client subsets. The fixed subsets satisfy

$$U_\ell \cap U_r = \emptyset, \quad \ell \neq r, \quad \bigcup_{\ell=1}^{N_{\text{sub}}} U_\ell = [N], \quad 3 \leq |U_\ell| \leq 5. \quad (3.8)$$

Let Π_t denote the client partition used in communication round t . FCPS requires

$$\Pi_t = \Pi \quad (3.9)$$

for every communication round t . Therefore, each client remains in the same fixed subset throughout collaborative training. The fixed subsets are used to construct the aggregation function vectors $\mathbf{y}_{U_\ell}$. The corresponding functional keys are generated once before collaborative training and reused throughout all communication rounds.

The proposed privacy-preserving FL system consists of the following three core entities:

- **Trusted authority (TA):** The TA acts as a trusted entity responsible for key management in the proposed cryptographic protocol. It does not participate in model training or gradient aggregation. It distributes initial keys. To aggregate a subset U_ℓ , the TA issues a functional key parameterized by an aggregation function vector $\mathbf{y}_{U_\ell} = (y_{U_\ell,1}, y_{U_\ell,2}, \dots, y_{U_\ell,N})^\top \in \{0,1\}^N$, where $y_{U_\ell,i} = 1$ for $i \in U_\ell$ and $y_{U_\ell,i} = 0$ otherwise.
- **Clients:** Based on local data, each client computes its local gradient, applies labeled MCFE encryption under the round label ℓ_t using the encryption key allocated by the TA, and uploads the generated labeled ciphertext to the server.
- **Aggregation server (AS):** After collecting the ciphertexts, the AS utilizes the functional key provided by the TA to execute secure aggregation directly within the ciphertext domain. The AS only decrypts ciphertexts carrying the same round label ℓ_t . The AS recovers the authorized fixed-subset sums, which are used to update the shared output layer weight vector. These sums constitute the authorized outputs available to the AS in the security analysis.

Algorithm 1 describes the initialization and the secure aggregation procedure used in each communication round. The input layer parameters are updated locally and are neither encrypted nor aggregated. Section 3.2 analyzes the case where the input layer parameters are fixed.

Remark 3.1. Compared with the ideal global gradient aggregation defined in (2.4), MCFE secure aggregation introduces additional computational overhead. In Algorithm 1, this overhead mainly comes from the encryption of $\mathbf{x}_{i,t}$ by clients, the generation of $dk_{y_{U_\ell}}$ by the TA, and the decryption of ciphertexts by the AS. Consequently, the additional time consumed by MCFE aggregation represents the computational overhead of computing the authorized fixed-subset sums from encrypted client gradients.

Algorithm 1 MCFE-based secure gradient aggregation algorithm.

• Step 0: Local computation and quantization (Clients)

- 1) Receive the shared output layer parameters w_t from the AS;
- 2) Update the input layer parameters ϕ_i^t using the local dataset;
- 3) Compute the output layer gradient $g_{i,t}$;
- 4) Encode $g_{i,t}$ into the non-negative integer plaintext vector $\mathbf{x}_{i,t}$ according to (3.1)–(3.2).

• Step 1: Key generation (TA)

- 1) Execute $\text{Setup}(1^\kappa, 1^d, 1^N)$ to generate master keys $\text{mpk}, \text{msk}, \{sk_i\}_{i \in [N]}$ and distribute encryption key sk_i to each client;
- 2) Partition clients into disjoint subsets U_ℓ (via FCPS). For each subset, construct a function vector $\mathbf{y}_{U_\ell}$;
- 3) Execute $\text{KeyGen}(\text{msk}, \mathbf{y}_{U_\ell})$ to generate the specific functional key $dk_{\mathbf{y}_{U_\ell}}$ for the AS.

• Step 2: Encryption and transmission (Clients)

- 1) Execute $\text{Encrypt}(sk_i, i, \ell_t, \mathbf{x}_{i,t})$ to encrypt the plaintext vector into the labeled ciphertext $ct_{i,t}^{\ell_t}$ and upload it to the AS.

• Step 3: Secure aggregation and global update (AS)

- 1) For each subset U_ℓ , execute $\text{Decrypt}(dk_{\mathbf{y}_{U_\ell}}, \ell_t, \{ct_{i,t}^{\ell_t}\}_{i \in [N]})$ directly in the ciphertext domain to obtain $X_{U_\ell,t}$;
 - 2) Recover the subset gradient sum vector $G_{U_\ell,t}$ from $X_{U_\ell,t}$ according to (3.5);
 - 3) Aggregate all subset sums and divide by total clients N to compute the recovered global average gradient $\tilde{g}_t = \frac{1}{N} \sum_{\ell=1}^{N_{\text{sub}}} G_{U_\ell,t}$;
 - 4) Update the output layer weight vector: $w_{t+1} = w_t - \eta_t \tilde{g}_t$.
-

3.2. Convergence analysis

During actual training, the input layer parameters $\{\phi_i^t\}_{i=1}^N$ are updated locally and affect the output layer gradients. Therefore, the complete training process involves coupled updates of the input and output layer parameters. For the conditional convergence analysis, we fix the input layer parameters during the analyzed output layer optimization process and denote their fixed values by $\{\phi_i\}_{i=1}^N$ for all communication rounds considered in the analysis. Accordingly, the local loss function in (2.1) and the global objective function in (2.2) reduce to

$$F_i(w \mid \phi_i) = \frac{1}{|\mathcal{D}_i|} \sum_{k=1}^{|\mathcal{D}_i|} (\tau_{i,k} - w^\top h_{i,k})^2 \quad (3.10)$$

and

$$F(w \mid \{\phi_i\}) = \frac{1}{N} \sum_{i=1}^N F_i(w \mid \phi_i), \quad (3.11)$$

respectively. Thus, $F(w \mid \{\phi_i\})$ remains fixed across communication rounds. Here, ϕ_i denotes the fixed value of ϕ_i^t in the analyzed problem. In practice, the recovered global average gradient $\tilde{g}_t$ computed

in Algorithm 1 contains the quantization error induced by the encoding–decoding process. Therefore, for convergence analysis, the same quantity is decomposed as $\tilde{g}_t = g_t + \vartheta_t$, where g_t is the ideal global average gradient defined in (2.4), and ϑ_t denotes the quantization error introduced by the encoding–decoding process. Instead of the ideal update in (2.5), we analyze the noisy update model

$$w_{t+1} = w_t - \eta_t \tilde{g}_t. \quad (3.12)$$

Assumption 3.2. For fixed $\{\phi_i\}_{i=1}^N$, the global conditional objective function $F(w \mid \{\phi_i\})$ is L -smooth with respect to w .

Lemma 3.3 ([24]). Under Assumption 3.2, for any two shared output layer weight vectors p and q , the global conditional objective function satisfies

$$F(q \mid \{\phi_i\}) \leq F(p \mid \{\phi_i\}) + \langle \nabla_p F(p \mid \{\phi_i\}), q - p \rangle + \frac{L}{2} \|q - p\|^2. \quad (3.13)$$

Because each client computes the full local gradient over its entire local dataset, the ideal global average gradient g_t defined in (2.4) is deterministic. Under the fixed input layer parameters setting defined in (3.10)–(3.11), it satisfies

$$g_t = \nabla_w F(w \mid \{\phi_i\}) \Big|_{w=w_t}. \quad (3.14)$$

Assumption 3.4. There exists a non-negative constant B such that the ideal global average gradient g_t satisfies

$$\|g_t\|^2 \leq B^2. \quad (3.15)$$

Assumption 3.5. The learning rate sequence $\{\eta_t\}$ satisfies

$$\sum_{t=0}^{\infty} \eta_t = \infty, \quad \sum_{t=0}^{\infty} \eta_t^2 < \infty. \quad (3.16)$$

Assumption 3.6. For all clients, communication rounds, and gradient dimensions, the scaled gradients do not exceed the clipping range, that is,

$$|g_{i,t,j}|S \leq Q_{\max}. \quad (3.17)$$

Therefore, the clipping operation is inactive, and the quantization error is only caused by nearest-integer rounding.

Under Assumption 3.6, for a uniform quantizer with nearest-integer rounding and scaling factor S , the per-dimension rounding error is bounded by $\frac{1}{2S}$. Hence, for a d -dimensional quantization error vector, the worst-case squared L_2 norm is the sum of the squared errors across all dimensions. Thus, it satisfies

$$\|\vartheta_t\|^2 = \sum_{j=1}^d (\vartheta_{t,j})^2 \leq \sum_{j=1}^d \left(\frac{1}{2S} \right)^2 = \frac{d}{4S^2}. \quad (3.18)$$

Remark 3.7. The bound in (3.18) applies to the case where the clipping operation does not change the scaled gradients. In this case, the quantization error comes only from nearest-integer rounding and is controlled by the scaling factor S . The clipping operation is still retained in the encoder to prevent scaled gradients from exceeding the clipping range $[-Q_{\max}, Q_{\max}]$.

Theorem 3.8. Under Assumptions 3.2 and 3.4–3.6, letting T represent the total number of communication rounds in the FL system, the sequence $\{w_t\}$ generated by (3.12) satisfies

$$\lim_{T \rightarrow +\infty} \sup \frac{\sum_{t=0}^{T-1} \eta_t \left\| \nabla_w F(w_t | \{\phi_i\}) \Big|_{w=w_t} \right\|^2}{\sum_{t=0}^{T-1} \eta_t} \leq O\left(\frac{d}{S^2}\right). \quad (3.19)$$

Proof. Rewrite $F(w_t) = F(w_t | \{\phi_i\}) \Big|_{w=w_t}$, $\nabla F(w_t) = \nabla_w F(w_t | \{\phi_i\}) \Big|_{w=w_t}$ for brevity. Applying Lemma 3.3 with $p = w_t$ and $q = w_{t+1} = w_t - \eta_t \tilde{g}_t$, we have

$$F(w_{t+1}) \leq F(w_t) - \eta_t \langle \nabla F(w_t), \tilde{g}_t \rangle + \frac{L\eta_t^2}{2} \|\tilde{g}_t\|^2. \quad (3.20)$$

Substituting $\tilde{g}_t = g_t + \vartheta_t$ into (3.20), we obtain

$$F(w_{t+1}) \leq F(w_t) - \eta_t \langle \nabla F(w_t), g_t \rangle - \eta_t \langle \nabla F(w_t), \vartheta_t \rangle + \frac{L\eta_t^2}{2} \|g_t + \vartheta_t\|^2. \quad (3.21)$$

Because $g_t = \nabla F(w_t)$, we have

$$\langle \nabla F(w_t), g_t \rangle = \|\nabla F(w_t)\|^2. \quad (3.22)$$

Applying Young's inequality $-\langle a, b \rangle \leq \frac{1}{2}\|a\|^2 + \frac{1}{2}\|b\|^2$ [25] to the cross term involving the quantization error ϑ_t , we obtain

$$-\langle \nabla F(w_t), \vartheta_t \rangle \leq \frac{1}{2} \|\nabla F(w_t)\|^2 + \frac{1}{2} \|\vartheta_t\|^2. \quad (3.23)$$

Moreover, using $\|a + b\|^2 \leq 2\|a\|^2 + 2\|b\|^2$ and Assumption 3.4,

$$\|g_t + \vartheta_t\|^2 \leq 2\|g_t\|^2 + 2\|\vartheta_t\|^2 \leq 2B^2 + 2\|\vartheta_t\|^2. \quad (3.24)$$

Substituting (3.22)–(3.24) into (3.21) yields the one-step bound

$$F(w_{t+1}) \leq F(w_t) - \frac{\eta_t}{2} \|\nabla F(w_t)\|^2 + \frac{\eta_t}{2} \|\vartheta_t\|^2 + L\eta_t^2 (B^2 + \|\vartheta_t\|^2). \quad (3.25)$$

Summing (3.25) over $t = 0, \dots, T-1$ and collecting the accumulated differences in the objective values yields

$$\frac{1}{2} \sum_{t=0}^{T-1} \eta_t \|\nabla F(w_t)\|^2 \leq F(w_0) - F(w_T) + \frac{1}{2} \sum_{t=0}^{T-1} \eta_t \|\vartheta_t\|^2 + L \sum_{t=0}^{T-1} \eta_t^2 (B^2 + \|\vartheta_t\|^2). \quad (3.26)$$

Because F is lower bounded by its global minimum F^* , we have $F(w_0) - F(w_T) \leq F(w_0) - F^*$. Dividing both sides of (3.26) by $\sum_{t=0}^{T-1} \eta_t$ and letting $T \rightarrow \infty$, Assumption 3.5 ensures that the term proportional to $\sum_{t=0}^{T-1} \eta_t^2$ vanishes after normalization by $\sum_{t=0}^{T-1} \eta_t$, yielding

$$\lim_{T \rightarrow +\infty} \sup \frac{\sum_{t=0}^{T-1} \eta_t \|\nabla F(w_t)\|^2}{\sum_{t=0}^{T-1} \eta_t} \leq \lim_{T \rightarrow +\infty} \sup \frac{\sum_{t=0}^{T-1} \eta_t \|\vartheta_t\|^2}{\sum_{t=0}^{T-1} \eta_t}. \quad (3.27)$$

Finally, applying the bound in (3.18) gives

$$\lim_{T \rightarrow +\infty} \sup \frac{\sum_{t=0}^{T-1} \eta_t \|\nabla F(w_t)\|^2}{\sum_{t=0}^{T-1} \eta_t} \leq O\left(\frac{d}{S^2}\right). \quad (3.28)$$

□

3.3. Security analysis

We first analyze the modulus requirement that prevents overflow in the quantized aggregation mechanism. Then, we establish the single-round ciphertext indistinguishability of the proposed protocol by reduction to the single-label indistinguishability (IND) security of the underlying labeled MCFE scheme. The admissible encoded-gradient collections agree on every authorized fixed-subset output; hence, the result addresses the ciphertext information not revealed by these outputs. In addition, the labeled MCFE syntax binds each ciphertext to its communication round. In round t , all valid ciphertexts are generated under the public label $\ell_t = t$. Therefore, the AS can only decrypt ciphertexts carrying the same label. For $t \neq t'$, a mixed ciphertext collection such as

$$\{ct_{1,t}^{\ell_t}, ct_{2,t}^{\ell_t}, ct_{3,t'}^{\ell_{t'}}\} \quad (3.29)$$

is not a valid input to the decryption algorithm and returns $\perp$. Hence, ciphertexts generated under different round labels cannot be combined as one valid input to the decryption algorithm. Nevertheless, the AS can separately obtain the authorized fixed subset outputs in different communication rounds and compare their changes over time.

Theorem 3.9. *For an authorized subset U_ℓ , recall that P denotes the plaintext modulus, and $Q_{\max}$ denotes the quantization bound (i.e., $|q_{i,t,j}| \leq Q_{\max}$). If the following condition is satisfied:*

$$P > |U_\ell|(Q_{\max} + K) \quad (3.30)$$

then the subset aggregated result can be recovered without modular overflow.

Proof. According to the FCPS, the secure aggregation and decryption are performed per subset. The encoded subset-sum component $X_{U_\ell,t,j}$ is given by

$$X_{U_\ell,t,j} = \sum_{i \in U_\ell} x_{i,t,j} = \sum_{i \in U_\ell} (q_{i,t,j} + K) = \sum_{i \in U_\ell} q_{i,t,j} + |U_\ell| \cdot K. \quad (3.31)$$

To recover the quantized sum without modular overflow, $X_{U_\ell,t,j}$ must be smaller than the modulus P . Because $q_{i,t,j} \leq Q_{\max}$, the encoded subset-sum component is bounded by

$$X_{U_\ell,t,j} \leq |U_\ell| \cdot (Q_{\max} + K). \quad (3.32)$$

Therefore, if $P > |U_\ell| \cdot (Q_{\max} + K)$, then $P > X_{U_\ell,t,j}$. In this case, the aggregation operation over the modular ring $\mathbb{Z}_P$ is equivalent to the corresponding integer-domain summation over $\mathbb{Z}$ for the considered subset and dimension. Hence, the encoded subset-sum component can be recovered without modular overflow. $\square$

Building upon the correct execution of the MCFE protocol, we analyze the confidentiality of the ciphertext collection generated in an arbitrary communication round t under the honest-but-curious AS model. The AS follows the protocol but attempts to distinguish the encrypted client inputs. We assume that the TA does not collude with the AS and only issues functional keys for the authorized fixed subsets in Π .

For an arbitrary communication round t , consider two collections of encoded client gradients

$$\mathcal{X}_t^{(b)} = \{\mathbf{x}_{i,t}^{(b)}\}_{i \in [N]}, \quad b \in \{0, 1\}. \quad (3.33)$$

The two collections are admissible challenge inputs if they produce identical outputs for every authorized fixed subset $U_\ell \in \Pi$, that is,

$$\sum_{i=1}^N y_{U_\ell, i} \mathbf{x}_{i,t}^{(0)} = \sum_{i=1}^N y_{U_\ell, i} \mathbf{x}_{i,t}^{(1)} \pmod{P}, \quad \ell = 1, 2, \dots, N_{\text{sub}}. \quad (3.34)$$

Equivalently,

$$\sum_{i \in U_\ell} \mathbf{x}_{i,t}^{(0)} = \sum_{i \in U_\ell} \mathbf{x}_{i,t}^{(1)} \pmod{P}. \quad (3.35)$$

Theorem 3.10. *Assume that the adopted labeled MCFE scheme satisfies static single-label IND security. If the two encoded-gradient collections in (3.33) satisfy (3.34), then any probabilistic polynomial-time adversary $\mathcal{A}$ has only negligible advantage in distinguishing their ciphertext collections under the round label ℓ_t .*

Proof. The challenger selects a random bit $b \in \{0, 1\}$ and generates

$$ct_{i,t}^{\ell_t, b} \leftarrow \text{Encrypt}(sk_i, i, \ell_t, \mathbf{x}_{i,t}^{(b)}), \quad i \in [N]. \quad (3.36)$$

Let

$$C_t^{(b)} = \{ct_{i,t}^{\ell_t, b}\}_{i \in [N]} \quad (3.37)$$

denote the resulting challenge ciphertext collection. After receiving mpk , ℓ_t , the fixed partition Π , the authorized functional keys $\{dk_{y_{U_\ell}}\}_{\ell=1}^{N_{\text{sub}}}$, and $C_t^{(b)}$, the adversary outputs a guess b' . Its advantage is defined as

$$\text{Adv}_{\text{prot}, \mathcal{A}}^{\text{IND}_1}(\kappa) = \left| \Pr[\mathcal{A}(C_t^{(0)}) = 1] - \Pr[\mathcal{A}(C_t^{(1)}) = 1] \right|. \quad (3.38)$$

Suppose that $\mathcal{A}$ achieves a non-negligible advantage. We construct an adversary $\mathcal{B}$ against the single-label IND security of the underlying labeled MCFE scheme. The adversary $\mathcal{B}$ submits $(\mathbf{x}_{i,t}^{(0)}, \mathbf{x}_{i,t}^{(1)})$ as the left and right challenge inputs for each client i under the label ℓ_t and requests only the functional keys associated with the authorized fixed subsets in Π .

Condition (3.34) guarantees that the two challenge collections produce identical outputs under every requested functional key. Hence, the challenge is valid in the single-label IND experiment. The MCFE challenger returns the ciphertext collection $C_t^{(b)}$, which $\mathcal{B}$ forwards to $\mathcal{A}$ together with the identical public information and authorized functional keys. The view provided to $\mathcal{A}$ has the same distribution as that in the proposed protocol.

The adversary $\mathcal{B}$ outputs the guess produced by $\mathcal{A}$. Therefore,

$$\text{Adv}_{\text{prot}, \mathcal{A}}^{\text{IND}_1}(\kappa) = \text{Adv}_{\text{MCFE}, \mathcal{B}}^{\text{IND}_1}(\kappa). \quad (3.39)$$

A non-negligible advantage of $\mathcal{A}$ would thus contradict the static single-label IND security of the underlying labeled MCFE scheme. Consequently,

$$\text{Adv}_{\text{prot}, \mathcal{A}}^{\text{IND}_1}(\kappa) \leq \text{negl}(\kappa). \quad (3.40)$$

□

Theorem 3.10 shows that two ciphertext collections are computationally indistinguishable whenever their encoded-gradient collections yield the same output for every authorized fixed subset. The encoded subset sums $X_{U_\ell, t}$ remain available to the AS, and any information implied by these sums lies outside the theorem's confidentiality claim.

Corollary 3.11. *Suppose that each communication round uses the distinct label $\ell_t = t$, where $t = 0, 1, \dots, T - 1$. Consider two encoded-gradient histories,*

$$\mathcal{X}^{(b)} = \left\{ \mathcal{X}_t^{(b)} \right\}_{t=0}^{T-1}, \quad b \in \{0, 1\}, \quad (3.41)$$

such that for every communication round t and every fixed authorized subset $U_\ell \in \Pi$,

$$\sum_{i \in U_\ell} \mathbf{x}_{i,t}^{(0)} = \sum_{i \in U_\ell} \mathbf{x}_{i,t}^{(1)} \pmod{P}. \quad (3.42)$$

Then, any probabilistic polynomial-time adversary $\mathcal{A}$ has only negligible advantage in distinguishing the corresponding T -round ciphertext transcripts.

Proof. By Theorem 3.10, the proposed protocol satisfies ciphertext indistinguishability for one round label. According to the one-to-many-label transformation in Lemma 2.5 of Abdalla et al. [23], single-label IND security extends to multiple distinct labels, with

$$\text{Adv}_{\text{prot}, \mathcal{A}}^{\text{IND}_T}(\kappa) \leq T \text{Adv}_{\text{MCFE}, \mathcal{B}}^{\text{IND}_1}(\kappa). \quad (3.43)$$

Here, $q_{\text{Enc}} = T$ because each communication round uses one distinct label. Because T is fixed in the proposed protocol, multiplying a negligible single-label advantage by T still yields a negligible function,

$$\text{Adv}_{\text{prot}, \mathcal{A}}^{\text{IND}_T}(\kappa) \leq \text{negl}(\kappa). \quad (3.44)$$

□

Under the considered passive external-eavesdropper model, we assume that the eavesdropper obtains neither client encryption keys nor authorized functional keys and does not collude with the TA, the AS, or any client. Its view consists only of the public information and the intercepted ciphertext collections and is therefore a projection of the honest-but-curious AS's view. Consequently, for two encoded-gradient histories that produce identical outputs for every authorized fixed subset in every communication round, the corresponding ciphertext transcripts intercepted by the external eavesdropper are also computationally indistinguishable.

Remark 3.12. *Under FCPS, the partition remains fixed throughout all communication rounds, that is, $\Pi_t = \Pi$. Hence, the AS does not obtain authorized outputs over different overlapping client combinations, and no additional cross-subset equations are introduced through cross-round client regrouping. However, the AS can separately obtain $X_{U_\ell, t}$ for different communication rounds and compare their temporal changes. Accordingly, the single-round and multi-round IND results exclude additional ciphertext leakage beyond the authorized fixed-subset outputs. They do not preclude statistical inference from these outputs and auxiliary prior information.*

4. Simulation results

We evaluate the proposed method on a proprietary hot-rolled steel dataset for strip-crown prediction. The dataset contains 10,000 samples with 34-dimensional process features. To emulate an FL scenario, the samples are distributed among $N = 20$ clients. Before collaborative training, FCPS divides the clients once into $N_{\text{sub}} = 6$ fixed subsets of 3–5 clients, and the resulting partition remains unchanged throughout all $T = 50$ communication rounds. The FCPS weighting parameters are set to $\alpha = 0.4$, $\beta = 0.3$, and $\gamma = 0.3$. The dataset is split into 80% training samples and 20% testing samples. The prediction model is a three-layer ANN with a 34–12–1 architecture. The number of communication rounds is set to $T = 50$, the learning rate is updated according to $\eta_t = \frac{0.01}{1+0.05t}$, the quantization scaling factor is set to $S = 8192$, and the plaintext modulus is set to $P = 2^{31} - 1$. In addition to prediction performance, we record the runtime distribution to evaluate the computational overhead introduced by MCFE secure aggregation. The runtime is decomposed into local gradient computation, MCFE secure aggregation, model evaluation, and other overhead.

To quantitatively evaluate prediction performance, we adopt two metrics: root mean square error (RMSE) and the prediction accuracy within a $\pm 10 \mu\text{m}$ error tolerance $\text{Acc}_{\pm 10\mu\text{m}}$ [26]. Let $\mathcal{D}_i^{\text{test}}$ denote the test dataset on client i , and let $n = \sum_{i=1}^N |\mathcal{D}_i^{\text{test}}|$ denote the total number of test samples. These metrics are defined as

$$\begin{aligned} \text{RMSE} &= \sqrt{\frac{1}{n} \sum_{i=1}^N \sum_{k=1}^{|\mathcal{D}_i^{\text{test}}|} (\tau_{i,k} - \hat{\tau}_{i,k})^2}, \\ \text{Acc}_{\pm 10\mu\text{m}} &= \frac{100\%}{n} \sum_{i=1}^N \sum_{k=1}^{|\mathcal{D}_i^{\text{test}}|} \mathbf{1}(|\tau_{i,k} - \hat{\tau}_{i,k}| \leq 10 \mu\text{m}), \end{aligned} \quad (4.1)$$

where $\mathbf{1}(\cdot)$ is the indicator function, which equals 1 when $|\tau_{i,k} - \hat{\tau}_{i,k}| \leq 10 \mu\text{m}$ and 0 otherwise. In the evaluation stage, $\hat{\tau}_{i,k}$ denotes the final predicted strip crown value of the k th test sample on client i .

We further define the quantization reconstruction error as

$$E_q(t) = \frac{1}{d} \|\tilde{g}_t - g_t\|^2. \quad (4.2)$$

4.1. Convergence and quantization analysis

We first validate the learning capability of the proposed scheme. Figure 1(a) shows that the test RMSE decreases rapidly in the early communication rounds and then approaches a stable level, showing empirical stabilization of the test RMSE under the evaluated setting. Figure 1(b) shows that $\text{Acc}_{\pm 10\mu\text{m}}$ increases during training and remains at a high level in later communication rounds. Figure 1(c) further shows that the quantization reconstruction error remains low during training. The prediction results in Figure 1(d) are distributed close to the ideal line, showing that the trained model can capture the relationship between rolling process parameters and strip crown. Table 2 shows that the runtime is mainly dominated by MCFE secure aggregation.

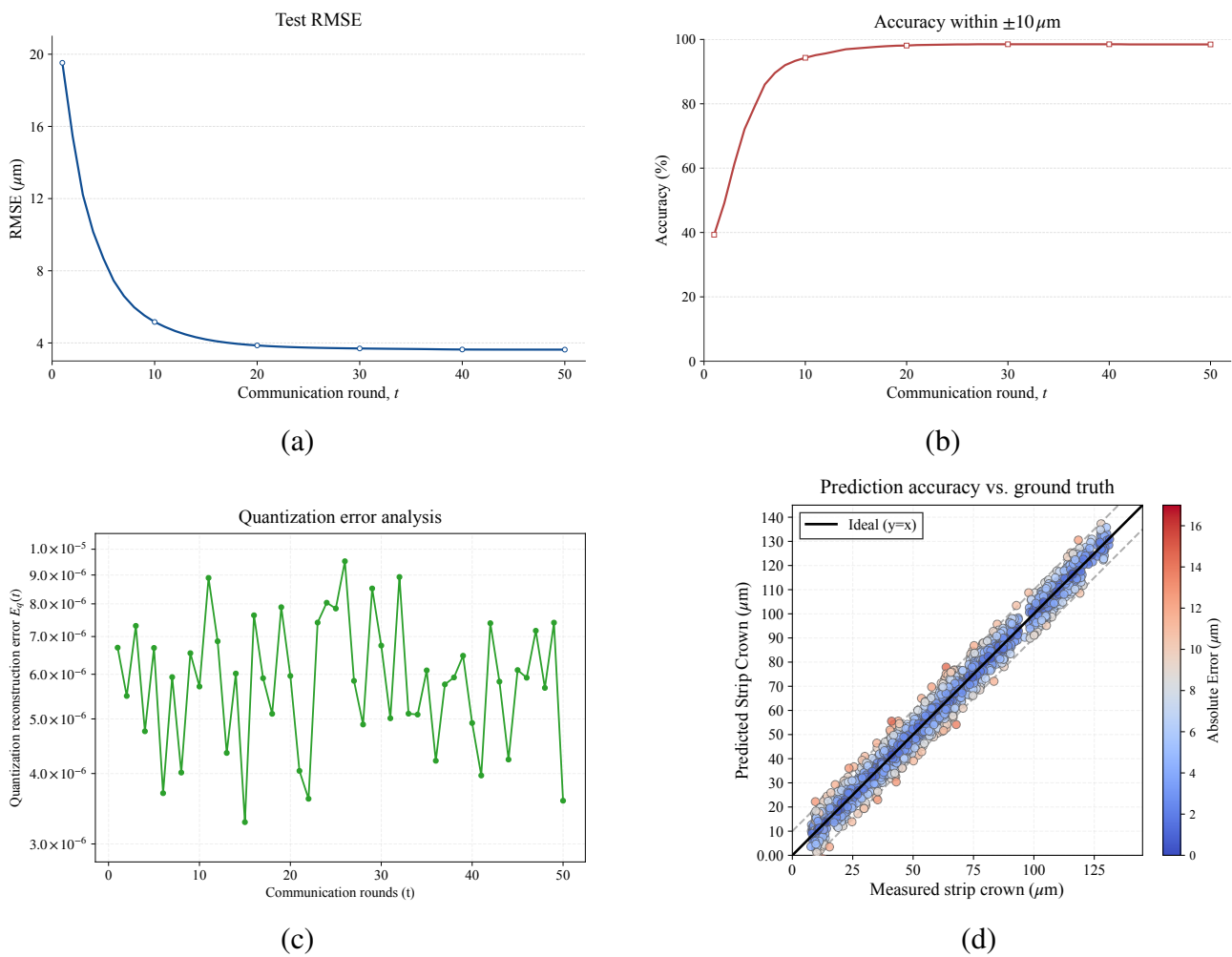


Figure 1. Encrypted experiment results: (a) Test RMSE convergence; (b) $\text{Acc}_{\pm 10\mu\text{m}}$ convergence; (c) Quantization reconstruction error; (d) Prediction accuracy vs. ground truth.

Table 2. Overall runtime distribution of the encrypted experiment.

Stage	Total time (s)	Mean time per round (s)	Share of total time
Local gradient computation	2.0704	0.0414	2.56%
MCFE secure aggregation	78.6280	1.5726	97.33%
Model evaluation	0.0689	0.0014	0.09%
Other overhead	0.0184	0.0004	0.02%
Total time	80.7858	1.6157	100.00%

4.2. Repeated experiments with multiple random seeds

To assess the variability of the proposed MCFE-based secure aggregation method, we repeated the experiment using five random seeds. For each seed, the data partition, model initialization, and fixed FCPS partition were regenerated, and all other experimental settings remain unchanged. The mean and

sample standard deviation are reported in Table 3. Across the five runs, the proposed method achieves a mean RMSE of $3.6290 \pm 0.5724 \mu\text{m}$, a mean $\text{Acc}_{\pm 10\mu\text{m}}$ of $97.20 \pm 0.42\%$, and a mean total time of $80.7858 \pm 2.7812 \text{ s}$.

Table 3. Repeated MCFE experiments over five random seeds.

Metric	Mean $\pm$ standard deviation
RMSE (μm)	3.6290 ± 0.5724
$\text{Acc}_{\pm 10\mu\text{m}}$ (%)	97.20 ± 0.42
Total time (s)	80.7858 ± 2.7812

4.3. Comparison with baseline methods

We further compare the proposed MCFE-based secure aggregation method with plaintext FL, Paillier HE, and local differential privacy (LDP) on the same 10,000-sample dataset. The same data partition, model architecture, and initialization are used for all methods. For the Paillier HE baseline, a standard additive homomorphic aggregation scheme with a 2048-bit key is adopted. The client-side output layer gradients are encoded using the same quantization scaling factor $S = 8192$ as the proposed method. For both Paillier HE and MCFE, the reported total time is measured using the same timing boundary and includes one-time cryptographic setup and key generation, gradient encoding, encryption, ciphertext aggregation, decryption, and decoding over all $T = 50$ communication rounds. For the LDP baseline, each client-side output layer gradient is clipped with an ℓ_2 -norm threshold of $C = 1$ and independently perturbed by Gaussian noise. Under replace-one adjacency at the client-update level and $\delta = 10^{-5}$, the target cumulative privacy budget ε is set to 3.0 over $T = 50$ communication rounds. The cumulative privacy loss is calculated using the additive composition property of Rényi differential privacy (RDP) for the Gaussian mechanism and is then converted to an (ε, δ) guarantee. Thus, the LDP baseline provides a cumulative $(3.0, 10^{-5})$ -LDP guarantee over the 50 communication rounds.

Table 4. Comparison with plaintext FL, Paillier HE, and LDP on the 10,000-sample dataset.

Method	RMSE (μm)	Total time (s)
Plaintext FL	3.5704	2.0662
MCFE	3.6290	80.7858
Paillier HE	3.5704	170.3833
LDP ($\varepsilon = 3.0$)	4.0440	2.2231

As shown in Table 4, using plaintext FL as the reference, the proposed MCFE-based secure aggregation method and Paillier HE achieve favorable prediction performance, with RMSE values of $3.6290 \mu\text{m}$ and $3.5704 \mu\text{m}$, respectively. The LDP baseline yields a slightly higher RMSE of $4.0440 \mu\text{m}$, indicating slightly lower prediction performance due to gradient clipping and Gaussian perturbation. In terms of runtime, MCFE requires 80.7858 s, which is substantially lower than the 170.3833 s required by Paillier HE but higher than those of plaintext FL and LDP.

4.4. Abnormal-gradient clipping experiment

We further investigate the effect of out-of-range abnormal output layer gradients. Among the 20 clients, 0–4 clients are selected as abnormal clients. Before fixed-point encoding, each output layer gradient component of an abnormal client is replaced by a wrong-direction value with an absolute magnitude of 5.0. Because $Q_{\max}/S = 32767/8192 \approx 3.9999$, these abnormal gradients exceed the encoding interval and trigger clipping. The results are reported in Table 5.

Table 5. Effect of abnormal client gradients on prediction performance.

Number of abnormal clients	Ratio of abnormal clients	RMSE (μm)	Acc $_{\pm 10\mu\text{m}}$ (%)
0	0%	3.6290	97.20
1	5%	4.8294	94.80
2	10%	6.6050	89.60
3	15%	8.6823	80.40
4	20%	12.5486	64.55

As the number and ratio of abnormal clients increase, RMSE increases, whereas Acc $_{\pm 10\mu\text{m}}$ decreases. These results show that clipping restricts abnormal gradients to the valid encoding interval, but the remaining abnormal updates still degrade prediction performance.

4.5. Sample size sensitivity analysis

We further examine the sensitivity of the proposed scheme to sample size. All experiments in this subsection use the same configuration as the main experiment, and only the investigated parameter is changed. Figure 2 shows the convergence behavior under different sample sizes. The effect of sample size on prediction performance is reported in Table 6.

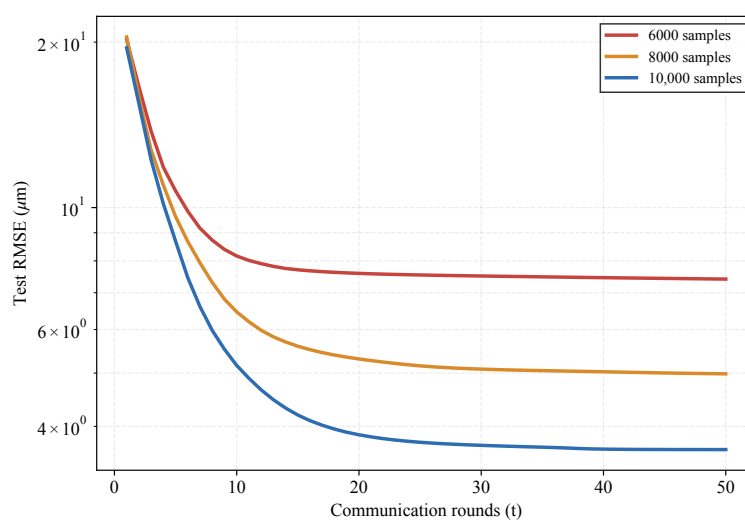


Figure 2. RMSE convergence under different sample sizes.

Table 6. Effect of sample size on prediction performance.

Total samples	Training samples	RMSE (μm)	Acc $_{\pm 10\mu\text{m}}$ (%)
6000	4800	7.4121	93.33
8000	6400	4.9836	96.75
10,000	8000	3.6290	97.20

The results show that a larger sample size improves both convergence behavior and final prediction accuracy. When the total number of samples increases from 6000 to 10,000, the final RMSE decreases from 7.4121 μm to 3.6290 μm .

The effects of non-identical and independently distributed (IID) data distributions across steel mills, fixed subset size, and quantization scaling factor have not been systematically evaluated. When the clipping operation is inactive, the per-dimension rounding error is bounded by $1/(2S)$. Increasing S reduces this bound but narrows the representable gradient range $[-Q_{\max}/S, Q_{\max}/S]$ for fixed $Q_{\max}$ and may increase clipping. Expanding the clipping range $[-Q_{\max}, Q_{\max}]$ requires reassessing the positive offset K and the modulus condition for the plaintext modulus P . The fixed subset size affects privacy granularity: Smaller subsets yield authorized fixed-subset outputs involving fewer clients, whereas larger subsets yield outputs aggregated over more clients. For a fixed total number of clients, the subset sizes also determine the number of authorized fixed subsets and hence affect the computational overhead of functional key generation and decryption. The reported configuration should therefore be regarded as an evaluated setting rather than a generally optimal choice.

5. Conclusions

This paper proposed a privacy-preserving FL framework for hot-rolled steel strip-crown prediction based on MCFE. The framework retains input layer parameters locally at clients and securely aggregates only the shared output layer gradients. By integrating gradient quantization, MCFE-based ciphertext-domain aggregation, and capacity-aware fixed client partitioning, the proposed scheme maintains the confidentiality of encrypted client gradients except for the information revealed by the authorized fixed-subset outputs during collaborative training.

Theoretical analysis showed that the encoded subset-sum can be correctly recovered under the specified modulus condition, and a conditional stationarity bound was derived for the shared output layer optimization phase under fixed local input layer parameters and quantization error. The security analysis further established single-round and multi-round ciphertext indistinguishability by reduction to the static single-label IND security of the underlying labeled MCFE scheme. The results showed that when two encoded-gradient histories yield the same output for every authorized fixed subset in every communication round, the AS has only negligible advantage in distinguishing the corresponding T -round ciphertext transcripts. Information contained in the authorized fixed-subset outputs remains outside this security claim. Experiments on a hot-rolled steel strip-crown prediction dataset achieved an RMSE of 3.6290 μm and an Acc $_{\pm 10\mu\text{m}}$ of 97.20% under the 10,000-sample setting, supporting the feasibility of MCFE-based privacy-preserving gradient aggregation for industrial strip-crown prediction. Future work will investigate the integration of differential privacy and data-tampering-resilient identification [27], as well as privacy-preserving state estimation based on the Cramér–Rao lower bound [28], into the proposed FL framework.

Use of AI tools declaration

The authors declare they have not used artificial intelligence (AI) tools in the creation of this article.

Acknowledgments

The work was supported by the National Natural Science Foundation of China under Grants 62433020 and T2293770.

Conflict of interest

The authors declare there are no conflicts of interest.

References

1. D. Matlock, J. Speer, Third generation of AHSS: Microstructure design concepts, in *Microstructure and Texture in Steels*, (2009), 185–205. https://doi.org/10.1007/978-1-84882-454-6_11
2. K. Muszka, D. Dziedzic, L. Madej, J. Majta, P. D. Hodgson, E. J. Palmiere, The development of ultrafine-grained hot rolling products using advanced thermomechanical processing, *Mater. Sci. Eng. A*, **610** (2014), 290–296. <https://doi.org/10.1016/j.msea.2014.05.051>
3. T. Cheng, G. Chen, Prediction of mechanical properties of hot-rolled strip steel based on PCA-GBDT method, *J. Phys. Conf. Ser.*, **1774** (2021), 012002. <https://doi.org/10.1088/1742-6596/1774/1/012002>
4. Y. Liu, X. Wang, J. Sun, G. Liu, H. Li, Y. Ji, Strip thickness and profile-flatness prediction in tandem hot rolling process using mechanism model-guided machine learning, *Steel Res. Int.*, **94** (2023), 2200447. <https://doi.org/10.1002/srin.202200447>
5. L. Song, D. Xu, X. Wang, Q. Yang, Y. Ji, Application of machine learning to predict and diagnose for hot-rolled strip crown, *Int. J. Adv. Manuf. Technol.*, **120** (2022), 881–890. <https://doi.org/10.1007/s00170-022-08825-w>
6. C. M. Sellars, Computer modelling of hot-working processes, *Mater. Sci. Technol.*, **1** (1985), 325–332. <https://doi.org/10.1179/mst.1985.1.4.325>
7. P. D. Hodgson, R. K. Gibbs, A mathematical model to predict the mechanical properties of hot rolled C-Mn and microalloyed steels, *ISIJ Int.*, **32** (1992), 1329–1338. <https://doi.org/10.2355/isijinternational.32.1329>
8. P. D. Hodgson, L. X. Kong, C. H. J. Davies, The prediction of the hot strength in steels with an integrated phenomenological and artificial neural network model, *J. Mater. Process. Technol.*, **87** (1999), 131–138. [https://doi.org/10.1016/S0924-0136\(98\)00344-6](https://doi.org/10.1016/S0924-0136(98)00344-6)
9. C. L. Mo, Q. Li, D. Z. Li, F. Feng, Z. D. Zhan, Prediction of the properties of hot rolling strip by using regression and neural network, *Acta Metall. Sin.*, **39** (2003), 1110–1114.

10. X. Li, Q. Jiang, X. Zhou, G. Cao, G. Wang, Z. Liu, Machine learning complex interactions among recovery, precipitation, and recrystallization for Nb micro-alloyed steels, *Met. Mater. Int.*, **30** (2024), 167–181. <https://doi.org/10.1007/s12540-023-01493-9>.
11. S. X. Zhao, S. C. Liu, Y. S. Jiang, B. Zhao, Y. L. Lv, J. Zhang, et al., Industrial foundation models (IFMs) for intelligent manufacturing: A systematic review, *J. Manuf. Syst.*, **82** (2025), 420–448. <https://doi.org/10.1016/j.jmsy.2025.06.011>
12. F. Roters, A. Aslam, Y. Bai, M. Büschelberger, K. Bulert, A. Butz, et al., Stahldigital: ontology-based workflows for the steel industry, *Adv. Eng. Mater.*, **27** (2025), 2402148. <https://doi.org/10.1002/adem.202402148>
13. T. Liu, J. Xia, G. Chen, J. H. Park, Synthesis of noisy sampled-data networked control systems under deception attacks: Direct current motor applications, *IEEE Internet Things J.*, **2026** (2026). <https://doi.org/10.1109/JIOT.2026.3680502>
14. Y. Fu, Z. Zhang, G. Zeng, X. Ren, Attack for range-only localization systems: A learning-based robust optimization approach, *IEEE Internet Things J.*, **2026** (2026). <https://doi.org/10.1109/JIOT.2026.3698447>
15. Z. Wang, J. Yang, Z. Liu, W. Ning, A two-layer deception attack detection framework for UAV, *IEEE Internet Things J.*, **13** (2026), 27228–27236. <https://doi.org/10.1109/JIOT.2026.3679816>
16. N. Haim, G. Vardi, G. Yehudai, O. Shamir, M. Irani, Reconstructing training data from trained neural networks, *Adv. Neural Inf. Process. Syst.*, **35** (2022), 22911–22924. <https://doi.org/10.52202/068431-1665>
17. A. Viand, P. Jattke, A. Hithnawi, SoK: Fully homomorphic encryption compilers, in *Proceedings of the 2021 IEEE Symposium on Security and Privacy (SP)*, San Francisco, CA, USA, (2021), 1092–1108. <https://doi.org/10.1109/SP40001.2021.00068>
18. K. Wei, J. Li, M. Ding, C. Ma, H. H. Yang, F. Farokhi, et al., Federated learning with differential privacy: Algorithms and performance analysis, *IEEE Trans. Inf. Forensics Secur.*, **15** (2020), 3454–3469. <https://doi.org/10.1109/TIFS.2020.2988575>
19. B. Knott, S. Venkataraman, A. Hannun, S. Sengupta, M. Ibrahim, L. van der Maaten, Crypten: Secure multi-party computation meets machine learning, *Adv. Neural Inf. Process. Syst.*, **34** (2021), 4961–4973.
20. X. Zhang, Y. Luo, T. Li, A review of research on secure aggregation for federated learning, *Future Internet*, **17** (2025), 308. <https://doi.org/10.3390/fi17070308>
21. R. Zhang, J. Han, L. Chen, Flexible threshold multi-client functional encryption for inner product in federated learning, preprint, arXiv:2510.15367. <https://doi.org/10.48550/arXiv.2510.15367>
22. J. Chotard, E. Dufour-Sans, R. Gay, D. H. Phan, D. Pointcheval, Decentralized multi-client functional encryption for inner product, in *International Conference on the Theory and Application of Cryptology and Information Security*, (2018), 703–732. https://doi.org/10.1007/978-3-030-03329-3_24
23. M. Abdalla, F. Benhamouda, R. Gay, From single-input to multi-client inner-product functional encryption, in *International Conference on the Theory and Application of Cryptology and Information Security*, (2019), 552–582. https://doi.org/10.1007/978-3-030-34618-8_19

24. L. Bottou, F. E. Curtis, J. Nocedal, Optimization methods for large-scale machine learning, *SIAM Rev.*, **60** (2018), 223–311. <https://doi.org/10.1137/16M1080173>
25. G. H. Hardy, J. E. Littlewood, G. Pólya, *Inequalities*, Cambridge University Press, Cambridge, 1952.
26. G. Li, D. Gong, X. Lu, D. Zhang, Ensemble learning based methods for crown prediction of hot-rolled strip, *ISIJ Int.*, **61** (2021), 1603–1613. <https://doi.org/10.2355/isijinternational.ISIJINT-2020-639>
27. J. Wang, J. Ke, J. Guo, Y. Zhao, Recursive binary identification with differential privacy and data tampering attacks, preprint, arXiv:2601.07608. <https://doi.org/10.48550/arXiv.2601.07608>
28. L. Guo, J. Wang, Y. Zhao, J. F. Zhang, State estimation with protecting exogenous inputs via Cramér-Rao lower bound approach, preprint, arXiv:2410.08756v2. <https://doi.org/10.1007/s11432-024-5021-y>



AIMS Press

© 2026 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0>)