eFL-Boost: Efficient Federated Learning for Gradient Boosting Decision Trees

Yamamoto, Fuki

Ozawa, Seiichi

Wang, Lihua

(Citation)
IEEE Access,10:43954-43963

(Issue Date)
2022-04-22

(Resource Type)
journal article

(Version)
Version of Record

IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received March 21, 2022, accepted April 5, 2022, date of publication April 22, 2022, date of current version April 29, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3169502

eFL-Boost: Efficient Federated Learning for Gradient Boosting Decision Trees

FUKI YAMAMOTO[1], **SEIICHI OZAWA**[2], **(Senior Member, IEEE), AND LIHUA WANG**[3]
[1]Department of Electrical and Electronic Engineering, Graduate School of Engineering, Kobe University, Kobe 657-8501, Japan
[2]Center for Mathematical and Data Sciences, Kobe University, Kobe 657-8501, Japan
[3]National Institute of Information and Communications Technology, Tokyo 184-8795, Japan

Corresponding author: Seiichi Ozawa (ozawasei@kobe-u.ac.jp)

This work was supported in part by Japan Science and Technology Agency (JST) CREST under Grant JPMJCR19F6, and in part by Japan Society for the Promotion of Science (JSPS) KAKENHI under Grant JP20K11826.

ABSTRACT Privacy protection has attracted increasing attention, and privacy concerns often prevent flexible data utilization. In most industries, data are distributed across multiple organizations due to privacy concerns. Federated learning (FL), which enables cross-organizational machine learning by communicating statistical information, is a state-of-the-art technology that is used to solve this problem. However, for gradient boosting decision tree (GBDT) in FL, balancing communication efficiency and security while maintaining sufficient accuracy remains an unresolved problem. In this paper, we propose an FL scheme for GBDT, i.e., efficient FL for GBDT (eFL-Boost), which minimizes accuracy loss, communication costs, and information leakage. The proposed scheme focuses on appropriate allocation of local computation (performed individually by each organization) and global computation (performed cooperatively by all organizations) when updating a model. It is known that tree structures incur high communication costs for global computation, whereas leaf weights do not require such costs and are expected to contribute relatively more to accuracy. Thus, in the proposed eFL-Boost, a tree structure is determined locally at one of the organizations, and leaf weights are calculated globally by aggregating the local gradients of all organizations. Specifically, eFL-Boost requires only three communications per update, and only statistical information that has low privacy risk is leaked to other organizations. Through performance evaluation on public data sets (ROC AUC, Log loss, and F1-score are used as metrics), the proposed eFL-Boost outperforms existing schemes that incur low communication costs and was comparable to a scheme that offers no privacy protection.

INDEX TERMS Machine learning, privacy-preserving, gradient boosting, federated learning, decision tree.

I. INTRODUCTION

Currently, future prediction and automation using data-driven artificial intelligence (AI) technologies are attracting increasing attention. Machine learning (ML) is a core technology for AI, which enables building statistical models based on various data to perform specific tasks. Gradient boosting decision tree (GBDT) [1] are popular ML models. A GBDT comprises multiple decision trees [2], which are trained by gradient boosting that sequentially builds weak learners to minimize the cost function. Significant GBDT implementations include XGBoost [3], LightGBM [4], and CatBoost [5], which have been used in various domains owing to their high learning efficiency and prediction performance [6]–[8].

Notably, concerns regarding privacy protection have been growing. These concerns led to related legal regulations, such as the European General Data Protection Regulation [9]. However, such regulations cannot completely eliminate the privacy risk and potential risks of malicious attacks, and greatly limit the flexibility of data utilization. Particularly, in various industries, data are scattered across multiple organizations, just like isolated data islands, preventing cross-organizational utilization. Therefore, privacy-preserving data mining [10] has attracted attention to realize secure and useful data utilization.

Conventional cross-organizational personal data utilization has been performed by a third-party provision of consented or

The associate editor coordinating the review of this manuscript and approving it for publication was Aysegul Ucar.

anonymized data [11]. However, this approach has concerns regarding the data loss and the information leakage, as well as security risks and inefficiencies caused by data aggregation. Therefore, there is a need to develop more efficient schemes for cross-organizational data utilization.

Federated learning (FL), proposed by Google [12], has been attracting attention as the state-of-the-art technology for cross-organizational data utilization [13], [14]. In FL, ML models are trained cooperatively by multiple organizations or users by aggregating statistical information on a third-party server. As FL requires only the communication of statistical information, it is expected to realize more efficient and secure cross-organizational data utilization than that using the conventional framework.

To address various real-world problems, FL schemes have been developed for various assumptions and ML models. For example, by focusing on the nature of data owned by organizations and users, Yang *et al.* [15] classified FL into horizontal FL, vertical FL, and federated transfer learning. Liu *et al.* [16] proposed federated forest, a vertical FL scheme, for random forests using cryptography. Hsu *et al.* [17] proposed an FL scheme for support vector machines and used it for Android malware detection. Phong *et al.* [18] proposed an FL scheme for deep neural networks using homomorphic encryption. Kim *et al.* [19] proposed a peer-to-peer FL scheme using blockchain that improves the vulnerability caused by the existence of a central server.

In GBDT, various methods that introduce FL and privacy-preserving techniques have been proposed. Cheng *et al.* [20] proposed SecureBoost, a vertical FL scheme, for GBDT, which uses secret computation with homomorphic encryption. Liu *et al.* [21] proposed FedXGB, which guarantees security against a central server and robustness against user dropout, using homomorphic encryption and secret sharing. Li *et al.* [22] proposed DPBoost, which guarantees differential privacy for GBDT, by defining accurate sensitivities and assigning noises efficiently.

Existing horizontal FL schemes for GBDT have limitations. Zhao *et al.* [23] proposed tree-based FL (TFL), an efficient scheme in which each organization takes turns updating the model; however, accuracy loss occurs on using this scheme. Yamamoto *et al.* [24] proposed federated GBDT with gradient-based model selection (F-GBDT-G), which improves the prediction performance to some extent by introducing a function into TFL that determines which owner to update the model. Li *et al.* [25] proposed similarity-based FL, which trains the model by leveraging the gradients of similar data from other organizations using locality sensitive hashing. Tian *et al.* [26] proposed private FL for GBDT (FederBoost), which can construct an exact tree by aggregating the gradient histograms from all data owners. In principle, the performance of FederBoost is the same as that of a single GDBT with the whole set of data provided by all organizations. However, this property causes several issues on high communication costs and large information leakage to other parties. Therefore, simultaneous achievement of good performance in terms of accuracy, security, and communication costs in FL remains an unresolved problem.

In this paper, we propose an FL scheme for GBDT, efficient FL for GBDT (eFL-Boost), which minimizes the accuracy loss, communication costs, and information leakage to other organizations. Here, we focus on the appropriate allocation of local computation (individually performed by each data owner) and global computation (cooperatively performed by all owners) to reduce communication costs and improve accuracy. Specifically, in eFL-Boost, a tree structure is locally determined by a single data owner, and leaf weights are globally calculated by aggregating the local gradients of all owners. Our primary contributions are summarized as follows.

- Globally computed GBDT has high prediction performance but high communication costs and information leakage. eFL-Boost achieves a trade-off between these costs and the accuracy by partially introducing local computation. However, introduction of local computation may cause accuracy loss. In the proposed eFL-Boost, this loss is compensated by the property of GBDT, which constructs a strong learner from multiple weak learners.
- To achieve the best trade-off between costs and accuracy, we attempt a balanced resource allocation in the local and global computations in the GBDT learning for the reduction of communication costs, information leakage, and accuracy loss. Specifically, in eFL-Boost, tree structures are determined locally not globally to avoid high communication costs and information leakage, while the leaf weight computation is conducted globally to avoid the degradation in the accuracy.

The remainder of this paper is organized as follows. Section II introduces several approaches to FL for multiple organizations, and Section III proposes a new approach called *eFL-Boost*. Section IV discuses the computational complexity, communication costs, and security assessment, and Section V gives the performance comparison using public data sets. Finally, Section VI summaries this paper and gives our future work.

II. FEDERATAED LEARNING FOR GBDT

A. GBDT

Herein, we briefly describe the basic learning algorithm in GBDT. Suppose that the t-th decision tree f_t in GBDT is constructed to minimize the following cost function $\mathcal{L}^t$ for N data,

$$\mathcal{L}^{(t)} = \sum_{i=1}^{N} L(y_i, \hat{y}_i + f_t(\mathbf{x}_i)) + \Omega(f_t), \tag{1}$$

where L is the loss function, $\mathbf{x}_i$ $(i = 1, \cdots, N)$ is the i-th feature vector, y_i is the i-th target, $\hat{y}_i$ is the prediction by the previous trees, and $\Omega(f_t)$ is the penalty term for the tree complexity. XGBoost and LightGBM use the following equation as the cost function, which is a quadratic approximation of the Taylor expansion around $\hat{y}_i$:

$$\mathcal{L}^{(t)} \approx \sum_{i=1}^{N} \left(L(y_i, \hat{y}_i) + g_i f_t(\mathbf{x}_i) + \frac{h_i f_t^{2}(\mathbf{x}_i)}{2} \right) + \Omega(f_t), \quad (2)$$

where $g_i = \partial_{\hat{y}_i} L\left(y_i, \hat{y}_i\right)$ and $h_i = \partial^2_{\hat{y}_i} L\left(y_i, \hat{y}_i\right)$. As the cost function is quadratic, the leaf weights $\hat{w}_j$ and split score for the node *score* can be expressed as follows [3]:

$$\hat{w}_j = -\frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda}, \quad (3)$$

$$score = \frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{G^2}{H + \lambda}, \quad (4)$$

where G_L and G_R (or H_L and H_R) denote the sum of the g (or h) of the data in the left and right nodes after splitting, respectively.

B. RELATED WORK

This paper focuses on the accuracy loss, communication costs, and information leakage to other owners. Herein, we compare eFL-Boost with three state-of-the-art FL schemes.

1) TFL [23]

Zhao *et al.* [23] proposed TFL, an FL scheme for GBDT, in which each data owner adds a local tree to the global model in turn. TFL only requires sending the global model to the next owner for each update, the communication costs and information leakage are low. To guarantee stronger security in TFL, a method to guarantee differential privacy is also proposed for models shared by all owners. By introducing this method, the risk of leakage of personal information from thresholds and leaf weights in the global model is reduced. However, from a practical perspective, differential privacy often comes at the cost of accuracy. To facilitate fair comparison, we adopt TFL without differential privacy in the following empirical study (V).

2) F-GBDT-G [24]

Yamamoto *et al.* [24] improved the prediction performance of TFL by introducing a function to actively select the next owner to update the model. In F-GBDT-G, the local trees built by each data owner are encrypted and aggregated to a central server. Then, the local tree having the highest learning loss is added to the global model. This model selection strategy is based on the so-called "greedy" approach; thus, the accuracy of the global model can be improved continuously by selecting the best local model. In terms of communication costs, F-GBDT-G requires only a single round of communication for each model update, thereby incurring low communication costs. Information leakage can be kept low because, other than the shared global model, the only information shared by the participants is a selection index, which is determined based on the gradient (g_i in (2)).

TABLE 1. Mathematical notations of variables.

Notations	Descriptions
n	#data provided by all owners
n_d	#data provided by the d-th owner
n_f	#input features
U	#model updates (= #trees)
l	tree depth
λ	regularization parameter
$\mathbf{X}$	data matrix
$\boldsymbol{y}$	target vector
$(\mathbf{X}_d, \boldsymbol{y}_d)$	a pair of data and target vectors of the d-th owner
$\boldsymbol{g}_d, \boldsymbol{h}_d$	gradient vectors computed by the d-th owner
$\mathbb{G}, \mathbb{H}$	accumulated gradient vectors at leaf nodes
$\boldsymbol{w}$	leaf weight vector
$\mathcal{D}$	data owners
$\mathcal{A}gg$	Aggregator
$\mathcal{B}$	Builder ($\mathcal{B} \in \mathcal{D}$)
$\mathbb{T}_i$	the i-th tree structure
$\mathcal{T}_i$	the i-th tree
$\{\mathcal{T}_1, \ldots, \mathcal{T}_i\}$	global model (Trees shared among owners)

3) FederBoost [26]

FederBoost is an FL scheme that can perform almost as well as a non-privacy-preserving GBDT scheme. However, compared with other schemes, FederBoost requires frequent communications and information leakage to other owners. Specifically, the histogram of the gradient corresponding to each node must be communicated several times depending on the depth of the tree. Each component of this histogram is the sum of the gradient values (g_i and h_i in (2)) of the data in each bin. Thus, from this histogram, data owners can obtain the gradient linked with the bins and the thresholds. As splitting proceeds, both the variance of the gradient and the amount of data in the node decrease. Thus, in the histogram of nodes close to the leaves, the target variable may be inferred from the gradient, and the target variable linked with the threshold and bins may indicate a minority group. By itself, the target variable rarely causes privacy risk; however, the target variable linked with other information may cause privacy leakage. To avoid inferring the target variable, a certain amount of data should be assigned to each bin of the histogram when training begins, which provides a security criterion for this scheme. However, during training, it is generally difficult to guarantee that this security criterion will be satisfied by all nodes.

III. EFFICIENT FL FOR GBDT

In TFL and F-GBDT-G, the global model is learned by simultaneously optimizing the local cost function of each owner; thus, bias in the local distribution can reduce accuracy. In contrast, in FederBoost, the global model is updated based on the global distribution; however, increased communication costs between data owners are inevitable when determining the decision tree's structure.

In this section, we propose eFL-Boost as a solution to the above-mentioned problems. Table 1 shows the list of mathematical notations of variables.

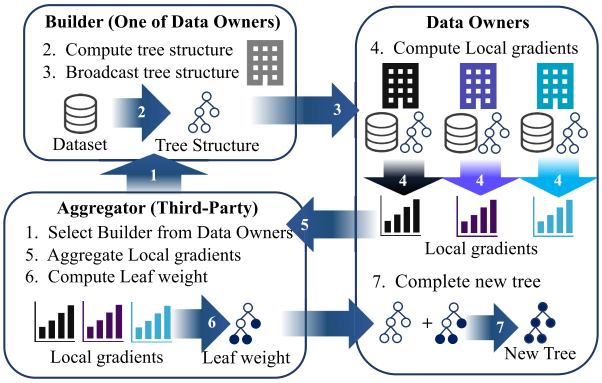
Builder (One of Data Owners)
2. Compute tree structure
3. Broadcast tree structure
2
Dataset
Tree Structure
3
1
Data Owners
4. Compute Local gradients
4
4
4
Local gradients
Aggregator (Third-Party)
1. Select Builder from Data Owners
5
5. Aggregate Local gradients
6. Compute Leaf weight
6
Local gradients
Leaf weight
7. Complete new tree
+
7
New Tree

FIGURE 1. Schematic of efficient federated learning for gradient boosting decision trees (eFL-Boost). It comprises Builder, Data Owners, and Aggregator. Builder is one of Data Owners, and is responsible for determining tree structures. Aggregator is responsible for aggregating gradients and calculating leaf weights. Data Owners compute a gradient vector based on the tree structure shared by Buider and sends it to Aggregator.

A. OVERVIEW

In this paper, we focus on appropriate allocation of global computation (performed cooperatively by all organizations) and local computation (performed individually by each organization) to reduce communication costs without compromising security and accuracy. A decision tree $\mathcal{T}$ can be divided into two parts: **tree structure** $\mathbb{T}$, which is a graph comprising multiple nodes with thresholds, and **leaf weights** $\boldsymbol{w}$, which are outputs of the model. In terms of communication costs, global tree structure determination requires tree depth-dependent costs; global leaf weight calculation requires only one round of communication. In terms of accuracy, leaf weights are considered to contribute more to the prediction performance because they are directly related to the output. Thus, in the proposed eFL-Boost, tree structure determination, which incurs large communication costs and provides a relatively small contribution to accuracy, is computed locally, and the leaf weights, which incur low communication costs and contribute significantly to accuracy, are computed globally.

A schematic of eFL-Boost is shown in Fig. 1. eFL-Boost operates under the assumption of horizontal FL schemes, where each data owner owns a dataset with a common feature set and different data sets. In addition, we assume following three participants in eFL-Boost.

Data Owners $\mathcal{D}$ are the organizations that own a dataset with a common feature set and different data sets. Here, each data owner $d \in \mathcal{D}$ computes $\mathbb{G}_d$ and $\mathbb{H}_d$, i.e., the sums of the gradients of data corresponding to each leaf, based on the tree structure shared by $\mathcal{B}$ and sends them to $\mathcal{A}gg$.

Builder $\mathcal{B}$ is an organization selected from $\mathcal{D}$ for each update. $\mathcal{B}$ is responsible for constructing the tree structure in addition to the role of $\mathcal{D}$.

Aggregator $\mathcal{A}gg$ is a third-party organization that does not have a dataset. $\mathcal{A}gg$ is responsible for aggregating the gradient vectors $\mathbb{G}_d$ and $\mathbb{H}_d$, calculating the leaf weights $\boldsymbol{w}$, and selecting $\mathcal{B}$. Here, $\mathcal{B}$ is selected based on a predetermined order, and the above three actions are executed simultaneously. Note that $\mathcal{A}gg$ does not belong to $\mathcal{D}$.

From a privacy protection perspective, $\mathcal{B}$ should not adopt a tree structure with leaves that correspond to only a few data points. Such a tree structure contains thresholds and leaf weights that are satisfied by only a small amount of data, which may increase the risk of personal information leakage. Therefore, a lower bound in the amount of data allocated to a leaf node should be properly predetermined so that the risk of personal information leakage cannot cause any practical issues.

B. ALGORITHM

Algorithm 1 presents the eFL-Boost algorithm. For illustration, the flow of the i-th update is as follows.

1) **Setup**. Each data owner $d \in \mathcal{D}$ updates the gradient of each data ($\boldsymbol{g}_d$ and $\boldsymbol{h}_d$) based on the previously constructed trees $\{\mathcal{T}_1, \ldots, \mathcal{T}_{i-1}\}$ (lines 4–6), and $\mathcal{A}gg$ selects $\mathcal{B}$ from $\mathcal{D}$ (line 3). In the proposed eFL-Boost, $\mathcal{A}gg$ selects a different $\mathcal{B}$ for each update to suppress bias in the global model.
2) **Local Tree Structure Determination.** $\mathcal{B}$ determines the tree structure $\mathbb{T}$ in (4) using its own dataset $\mathbf{X}_{\mathcal{B}}$ (line 7) and shares $\mathbb{T}$ with $\mathcal{D}$ (line 8).
   Note: Global computation by all owners of tree structures, as in FederBoost (II-B3), requires several communications of gradient histograms. In the proposed eFL-Boost, these costs are reduced by applying local computation. However, this local computation can reduce accuracy in each tree. But, this accuracy loss can be compensated to some extent, because GBDT can construct a strong learner from multiple weak learners.
3) **Aggregating Gradients for Each Leaf.** Each data owner $d \in \mathcal{D}$ computes $\mathbb{G}_d$ and $\mathbb{H}_d$, i.e., the sums of the gradients (g_i and h_i in (2)) corresponding to each leaf, with $\mathbb{T}$ and $\mathbf{X}_d$ (line 10) and sends them to $\mathcal{A}gg$ (line 11).
4) **Global Leaf Weight Calculation.** $\mathcal{A}gg$ aggregates $\mathbb{G}_d$ and $\mathbb{H}_d$ (line 13), calculates leaf weights $\boldsymbol{w}$ globally from $\mathbb{G}$ and $\mathbb{H}$ (line 14), and sends $\boldsymbol{w}$ to $\mathcal{D}$ (line 15).
   Note: Each tree computed locally, as in TFL (II-B1) and F-GBDT-G (II-B2), improves the cost function based on the local distribution, which can cause significant accuracy loss. However, in the proposed eFL-Boost, global leaf weight calculation by all owners enables cost function optimization based on the global distribution, which reduces this accuracy loss. In addition, only a single round of communication is required in this step (IV-A), and its contents involve low privacy risks (IV-C).
5) **Add a Completed Tree to the Global Model.** Each data owner $d \in \mathcal{D}$ combines $\mathbb{T}$ and $\boldsymbol{w}$ to obtain the i-th tree $\mathcal{T}_i$, which is then added to the global model (line 17).

With the proposed eFL-Boost, the computations of the tree structure and leaf weights are performed by different

```
Algorithm 1: Algorithm of eFL-Boost
  Input: $U$: Number of updates,
         $(\mathbf{X}_d, \boldsymbol{y}_d)$: Dataset for $d \in \mathcal{D}$,
         $\lambda$: Regularization parameter
  Output: $\{\mathcal{T}_1, \ldots, \mathcal{T}_U\}$: Global model
1  begin
2      for $i \leftarrow 1$ to $U$ do
3          $\mathcal{A}gg$ selects $\mathcal{B}$ from $\mathcal{D}$.
4          for each data owner $d \in \mathcal{D}$ do
5              $d$ updates gradients $\boldsymbol{g}_d$ and $\boldsymbol{h}_d$ based on $\mathcal{T}_{i-1}$
                 and $(\mathbf{X}_d, \boldsymbol{y}_d)$
6          end
7          $\mathcal{B}$ computes $\mathbb{T}_i$ from $\mathbf{X}_B, \boldsymbol{g}_B$ and $\boldsymbol{h}_B$.
8          $\mathcal{B}$ sends $\mathbb{T}_i$ to $\mathcal{D}$.
9          for each data owner $d \in \mathcal{D}$ do
10             $d$ computes $\mathbb{G}_d$ and $\mathbb{H}_d$ from $\mathbf{X}_d, \boldsymbol{g}_d$ and $\boldsymbol{h}_d$.
11             $d$ sends $\mathbb{G}_d$ and $\mathbb{H}_d$ to $\mathcal{A}gg$.
12         end
13         $\mathcal{A}gg$ computes $\mathbb{G} = \sum_{d \in \mathcal{D}} \mathbb{G}_d$ and
             $\mathbb{H} = \sum_{d \in \mathcal{D}} \mathbb{H}_d$.
14         $\mathcal{A}gg$ computes $\boldsymbol{w}_i = -\frac{\mathbb{G}}{\mathbb{H}+\lambda}$.
15         $\mathcal{A}gg$ sends $\boldsymbol{w}_i$ to $\mathcal{D}$.
16         for each data owner $d \in \mathcal{D}$ do
17             $d$ combines $\mathbb{T}_i$ and $\boldsymbol{w}_i$ to obtain $\mathcal{T}_i$.
18         end
19     end
20 end
```

organizations (**Local Tree Structure Determination** and **Global Leaf Weight Calculation**) to achieve an optimal trade-off between accuracy, communication costs, and security (i.e., information leakage). Several schemes that use only local and global computations have been proposed previously; however, to the best of our knowledge, no existing scheme employs a hybrid combination of local and global computation.

After all the updates are completed, each owner obtains a common set of complete decision trees. Therefore, in the inference stage, the model can be treated as a non-privacy-preserving scheme.

IV. DISCUSSIONS

In this section, we discuss the properties of eFL-Boost. In particular, we compare eFL-Boost with existing schemes in terms of communication costs, computational complexity, and information leakage.

A. COMMUNICATION COSTS

In this subsection, let us estimate the communication costs. The number of communications required by each scheme to update the global model is listed in Table 2, where l denotes the tree depth.

eFL-Boost requires three communications: the sharing of tree structures by $\mathcal{B}$ (Algorithm 1, line 8), transmission of gradient values by $\mathcal{D}$ (Algorithm 1, line 11), and transmission of leaf weights by $\mathcal{A}gg$ (Algorithm 1, line 15). Therefore, eFL-Boost requires $O(U)$ communication in the training phase with U iterations, similar to TFL [23] and F-GBDT-G [24]. Meanwhile, FederBoost [26] requires $O(lU)$ communication. Therefore, eFL-Boost can train a model with lower communication costs than FederBoost.

TABLE 2. Communication costs required by each scheme for each update, where *l* denotes the tree depth.

Schemes	Times
TFL [23]	1
F-GBDT-G [24]	2
FederBoost [26]	$2l$
eFL-Boost	3

TABLE 3. Computational complexity required by each scheme for each update.

Schemes	Learning	Prediction
Non-pp	$O(nn_f)$	$O(n)$
TFL [23]	$O(n_d n_f)$	$O(n)$
F-GBDT-G [24]	$O(n_d n_f)$	$O(n)$
FederBoost [26]	$O(n_d n_f)$	$O(n)$
eFL-Boost	$O(n_d n_f)$	$O(n)$

B. COMPUTATIONAL COMPLEXITY

In this subsection, we describe the computational complexity. The computational complexity required by each scheme for update and inference is listed in Table 3, where n_d is the number of data owned by the d-th owner, n is the total number of data provided by all owners, and n_f is the number of features.

As in the conventional GBDT, the computational complexity of non-privacy-preserving GBDT (Non-pp), which is trained on the collected data, is $O(nn_f)$. In TFL and F-GBDT-G, the computational complexity is $O(n_d n_f)$ because same algorithm as normal GBDT is used for each update. FederBoost requires the construction of the tree $O(n_d n_f)$ and the encryption of the histogram $O(n_d n_f)$; hence, the computational complexity is $O(n_d n_f)$ as well.

eFL-Boost requires three calculations: tree structure determination by $\mathcal{B}$ (Algorithm 1, line 7), summation of gradients of data in each leaf by $\mathcal{D}$ (Algorithm 1, line 10), and calculation of leaf weights by $\mathcal{A}gg$ (Algorithm 1, lines 13-14). Their computational complexities are $O(n_d n_f)$, $O(n_d)$, and $O(2^l)$ (generally $2^l \ll n_d$), respectively. Therefore, tree structure determination by $\mathcal{B}$ is the most dominant.

In addition, in all schemes, all owners share complete GBDT. Thus, the computational complexity required for inference is equal to that required for normal GBDT.

C. SECURITY ANALYSIS

In this subsection, we conduct the security analysis for the proposed eFL-Boost and the similar FL methods in II-B based on information leakage to other data owners. Let us assume that $\mathcal{D}$ and $\mathcal{A}gg$ attempt to obtain sensitive information from other data owners under the following conditions: (a) attackers do not use sophisticated techniques, (b) all

TABLE 4. Information obtained by data owners $\mathcal{D}$ and an aggregator $\mathcal{A}gg$ in each update by correctly executing each scheme, where $Hist$ is the gradient histogram, $\mathcal{T}_d$ is the decision tree of owner d, $Hist_d$ is the gradient histogram of owner d, and Sel is the selection index used in F-GBDT-G.

Schemes	$\mathcal{D}$	$\mathcal{A}gg$
TFL [23]	$\mathcal{T}$	-
F-GBDT-G [24]	$\mathcal{T}$	$\text{Enc}(\mathcal{T}_d), Sel$
FederBoost [26]	$\mathcal{T}, Hist$	$\text{Enc}(Hist_d)$
eFL-Boost	$\mathcal{T}$	$\mathbb{G}_d, \mathbb{H}_d$

participants use secure communication channels and the contents are not intercepted by others, and (c) all participants do not collude with each other.

Table 4 shows the information obtained by $\mathcal{D}$ and $\mathcal{A}gg$ through each scheme. In eFL-Boost, in addition to these participants, there is $\mathcal{B}$; however, because $\mathcal{B}$ is included in $\mathcal{D}$, $\mathcal{B}$ and $\mathcal{D}$ are treated as the same in the table. There is no entry for $\mathcal{A}gg$ in TFL because no $\mathcal{A}gg$ in TFL.

First, focusing on $\mathcal{A}gg$, only low-risk information is leaked in all schemes. In F-GBDT-G and FederBoost, the encrypted global model and histograms are leaked; however, no useful information can be obtained without the corresponding key. In the proposed eFL-Boost, $\mathcal{A}gg$ obtains $\mathbb{G}_d$ and $\mathbb{H}_d$, i.e., the sum of gradients for each leaf computed by each owner. However, $\mathcal{A}gg$ cannot associate each component of $\mathbb{G}_d$ and $\mathbb{H}_d$, i.e., the aggregated gradients, with other information because $\mathcal{A}gg$ does not have the tree structure.[1] Therefore, in eFL-Boost, $\mathcal{A}gg$ cannot obtain any useful information.

Second, focusing on $\mathcal{D}$, the proposed eFL-Boost has the lowest privacy risk in the existing schemes. Here, we separately consider the global model leaking to $\mathcal{D}$ in all schemes and the other information.

- *The global models in eFL-Boost and FederBoost have a lower privacy risk than those in TFL and F-GBDT-G.* The global models in eFL-Boost and FederBoost differ from those in TFL and F-GBDT-G, depending on whether each tree is based on a global or local distribution. Each tree in GBDT comprises a tree structure and leaf weights, where the leaf weights are calculated from the gradient of the loss function. In the leaf nodes, the target variable y may be inferred from the leaf weights when the variance of the gradients is reduced; thus, the global models leak the target variable y linked with the threshold to $\mathcal{D}$. If the amount of data corresponding to each leaf is small, the target variable y linked to the threshold indicates a minority group, which can result in privacy risks. From the above, a global model based on a global distribution (computed with more data) is considered to have lower privacy risks.
- *With the exception of the global model, eFL-Boost has lower privacy risks than FederBoost with respect to information leaked to $\mathcal{D}$.* In the proposed eFL-Boost, with the exception of the global model, no any other information is leaked to $\mathcal{D}$; however, in FederBoost, the gradient histograms corresponding to each node are leaked to $\mathcal{D}$ in addition to the global model. This histogram leaks information on the gradient linked with the threshold and bins to $\mathcal{D}$. In FederBoost, prior to training, this histogram is formed such that each bin contains several data points; however, as splitting proceeds, the variance of the gradients and the amount of data in each bin decrease. Thus, in the histogram of nodes close to the leaves, the target variable y may be inferred from the gradient, and the target variable y may indicate an individual or a minority group. In this case, the target variable y linked with thresholds or bins, which may indicate a minority group or an individual, is leaked to $\mathcal{D}$; thus, with FederBoost, privacy risks may occur in personal data analysis.

Note: In an experiment conducted to evaluate the proposed eFL-Boost, we set a lower limit on the amount of data to calculate leaf weights in order to avoid similar privacy leakage, and we tested its impact on accuracy (V-E).

TABLE 5. Data sets information. The tasks in the data sets below are binary classification.

Data set	#Data	#Features	Positive Data Ratio
Credit [29]	284805	30	0.2%
Breast [30]	569	30	62.7%
Biodeg [31]	1053	41	33.6%
German [32]	999	24	70.0%
Magic [33]	19018	10	64.8%

V. EXPERIMENTS

In this section, we carry out the performance comparisons with other related GBDT methods for the benchmark data sets in Table 5.

A. EXPERIMENTAL SETUPS

In this experiment, a personal computer with 64 [GB] RAM and Ubuntu 18.04 is used, and Python and Cython are used to implement the proposed eFL-Boost algorithm.

The popular benchmark datasets are used in the experiments (see Table 5). Here, only "Credit" is an imbalanced dataset. eFL-Boost works as the horizontal FL model; therefore, the datasets are divided horizontally into several subsets, each of which is allocated to an owner.

Here, 5-fold-cross-validation is performed with the three evaluation metrics: F1-score, area under a receiver operating characteristic curve (ROC-AUC), and logarithmic loss (Log loss). The reasons for selecting these metrics are as follows. F1-score is selected because it enables robust evaluation in this experiment involving an imbalanced dataset. ROC-AUC is selected because it enables robust performance evaluation independent of the threshold probability in classification. Log loss was selected because it is used as a loss function during training and can consider the range of error. In addition, the

[1] Although $\mathbb{G}_d$ and $\mathbb{H}_d$ have nearly no privacy risk even in plaintext, in the real world, organizations and users may resist providing this information. This problem can be solved by introducing secure computation using homomorphic encryption (e.g., [27], [28]). Concretely, $d \in \mathcal{D}$ encrypts $\mathbb{G}_d$ and $\mathbb{H}_d$ before sending them to $\mathcal{A}gg$ (line 11), and $\mathcal{A}gg$ computes $\mathbb{G}$ and $\mathbb{H}$ without decryption (line 13).

TABLE 6. Experimental results on the effect of the number of participating owners on prediction performance. Bold values in the table indicate the best in the column, except for FederBoost. The values with asterisks "*" and "" in the table denote the results of Welch's t-test (one-sided test). It is indicated that the performance of the proposed eFL-Boost statistically significantly outperforms that of the corresponding scheme, where "*" and "**" denote significance levels of 1% and 5%, respectively. The fractions in the table indicate the proportion of participating data owners, where the denominator represents the number of all data owners and the numerator represents the number of participating data owners. Each data owner has 10% of all data.**

Dataset	Scheme	F1-score			Log loss			AUC		
		3/10	5/10	10/10	3/10	5/10	10/10	3/10	5/10	10/10
Credit	FederBoost [26]	0.821	0.838	0.861	0.00411	0.00317	0.00258	0.974	0.974	0.980
	Individual	0.754*	0.745**	0.750**	0.00411*	0.00418*	0.00415**	0.965	0.965	0.966
	TFL [23]	0.797	0.797	0.814	0.00406	0.00364	0.00333*	0.968	0.974	0.974
	F-GBDT-G [24]	0.799	0.797	0.825	0.00399	0.00358	0.00358*	0.970	0.974	0.973
	eFL-Boost	**0.805**	**0.832**	**0.843**	**0.00377**	**0.00306**	**0.00265**	**0.971**	**0.977**	**0.980**
Breast	FederBoost [26]	0.955	0.965	0.976	0.151	0.118	0.102	0.988	0.991	0.992
	Individual	0.931	0.933	0.933*	0.217	0.209	0.210*	0.978	0.979	0.978
	TFL [23]	0.952	0.954	0.952	0.177	0.143	0.149	**0.984**	0.986	0.987
	F-GBDT-G [24]	0.954	0.955	0.952	0.173	0.151	0.155	**0.984**	0.987	0.987
	eFL-Boost	**0.956**	**0.959**	**0.963**	**0.164**	**0.130**	**0.117**	**0.984**	**0.988**	**0.989**
Biodeg	FederBoost [26]	0.740	0.751	0.772	0.466	0.414	0.421	0.888	0.902	0.911
	Individual	0.672	0.664*	0.671*	0.550	0.559	0.539*	0.846	0.844	0.849*
	TFL [23]	0.712	0.729	0.755	0.535	0.513	0.475*	0.876	0.888	0.896
	F-GBDT-G [24]	0.711	0.733	0.744	0.520	0.463	0.472	0.878	0.890	0.894
	eFL-Boost	**0.730**	**0.748**	**0.772**	**0.431**	**0.393**	**0.359**	**0.885**	**0.899**	**0.909**
German	FederBoost [26]	0.818	0.822	0.839	0.610	0.558	0.529	0.753	0.764	0.787
	Individual	0.812	0.811	0.811*	0.611	0.619	0.632*	0.720*	0.710*	0.711**
	TFL [23]	**0.823**	0.820	0.831	0.593	0.582	0.558	0.759	0.755	0.774
	F-GBDT-G [24]	**0.823**	0.820	0.830	0.569	0.590	0.558	0.760	0.756	0.771
	eFL-Boost	**0.823**	**0.825**	**0.839**	**0.540**	**0.544**	**0.504**	**0.762**	**0.761**	**0.787**
Magic	FederBoost [26]	0.902	0.905	0.910	0.322	0.312	0.296	0.922	0.927	0.934
	Individual	0.894**	0.893**	0.893**	0.348**	0.348**	0.347**	0.911**	0.911**	0.911**
	TFL [23]	0.897	0.900*	0.902**	0.341**	0.332*	0.322**	0.916*	0.919**	0.924**
	F-GBDT-G [24]	0.897*	0.900*	0.902**	0.343*	0.333**	0.322**	0.916*	0.919*	0.924*
	eFL-Boost	**0.901**	**0.904**	**0.908**	**0.325**	**0.313**	**0.303**	**0.922**	**0.926**	**0.931**

TABLE 7. Experimental results on the influence of the amount of data owned on the prediction performance. Bold values in the table indicate the best in the column, except for FederBoost. The values with asterisks "*" and "" in the table denote the results of Welch's t-test (one-sided test). It is indicated that the performance of the proposed eFL-Boost statistically significantly outperforms that of the corresponding scheme, where "*" and "**" denote significance levels of 1% and 5%, respectively. The fractions in the table indicate the proportion of participating data owners, where the denominator represents the number of all data owners and the numerator represents the number of participating data owners. Notably, the amount of data for each data owner increases as the total number of owners decreases. Specifically, each owner's data volume is 10%, 20% and 33% of total volume.**

Dataset	Scheme	F1-score			Log loss			AUC		
		3/10	3/5	3/3	3/10	3/5	3/3	3/10	3/5	3/3
Credit	FederBoost [26]	0.821	0.840	0.861	0.00411	0.00390	0.00258	0.974	0.974	0.980
	Individual	0.754*	0.801	0.821	0.00411*	0.00367	0.00329	0.966	0.969	0.974
	TFL [23]	0.797	0.827	0.847	0.00406	0.00340	0.00303	0.968	0.974	**0.979**
	F-GBDT-G [24]	0.799	0.824	0.844	0.00400	0.00328	0.00301	0.970	0.976	0.978
	eFL-Boost	**0.805**	**0.832**	**0.850**	**0.00377**	**0.00286**	**0.00275**	**0.971**	**0.978**	**0.979**
Breast	FederBoost [26]	0.955	0.971	0.976	0.151	0.107	0.102	0.988	0.992	0.992
	Individual	0.931	0.962	0.962	0.217	0.121	0.129	0.978	0.991	0.990
	TFL [23]	0.952	0.961	0.971	0.177	0.129	0.110	**0.984**	0.990	**0.992**
	F-GBDT-G [24]	0.954	0.963	0.970	0.173	0.130	0.106	**0.984**	0.990	**0.992**
	eFL-Boost	**0.956**	**0.966**	**0.973**	**0.164**	**0.111**	**0.101**	**0.984**	**0.991**	**0.992**
Biodeg	FederBoost [26]	0.740	0.754	0.772	0.466	0.438	0.421	0.888	0.901	0.911
	Individual	0.671	0.702	0.730	0.539	0.494	0.444	0.849	0.871	0.887
	TFL [23]	0.712	0.735	0.759	0.535	0.471	0.469	0.876	0.893	0.900
	F-GBDT-G [24]	0.711	0.743	0.771	0.520	0.474	0.438	0.878	0.895	0.904
	eFL-Boost	**0.730**	**0.747**	**0.776**	**0.431**	**0.413**	**0.393**	**0.885**	0.897	**0.907**
German	FederBoost [26]	0.818	0.822	0.839	0.610	0.617	0.529	0.753	0.766	0.787
	Individual	0.811	0.811*	0.822	0.632	0.625*	0.590*	0.720*	0.737*	0.760
	TFL [23]	**0.823**	0.819	0.833	0.593	0.610	0.520	0.759	0.758	0.786
	F-GBDT-G [24]	**0.823**	0.824	0.837	0.569	0.577	0.515	0.760	0.767	0.786
	eFL-Boost	**0.823**	**0.827**	**0.841**	**0.540**	**0.541**	**0.508**	**0.762**	**0.774**	**0.788**
Magic	FederBoost [26]	0.902	0.907	0.910	0.322	0.307	0.296	0.922	0.930	0.934
	Individual	0.894**	0.898**	0.903**	0.348**	0.331**	0.320*	0.911**	0.919**	0.924*
	TFL [23]	0.897	0.904	0.906	0.341**	0.315	0.307	0.916*	0.926	0.929
	F-GBDT-G [24]	0.897*	0.904	0.906	0.343*	0.315	0.305	0.916*	0.926	0.930
	eFL-Boost	**0.901**	**0.906**	**0.908**	**0.325**	**0.311**	**0.303**	**0.922**	**0.927**	**0.931**

statistical significance of the results was evaluated with the Welch's t-test (one-sided test) [34].

The following five GBDT schemes are selected for comparison. **Non-pp** indicates the case that the training of

TABLE 8. Experimental results on the influence of bias on the amount of data owned on the prediction performance. Bold values in the table indicate the best in the column. The values with asterisks "*" and "" in the table denote the results of Welch's t-test (one-sided test). It is indicated that the performance of the proposed eFL-Boost statistically significantly outperforms that of the corresponding scheme, where "*" and "**" denote significance levels of 1% and 5%, respectively. The ratios in the table show the ratios of the amount of data among the owners. eFL-Boost shows the best result in most cases.**

Dataset	Scheme	F1-score			Log loss			AUC		
		1:1:1	8:1:1	6:2:2	1:1:1	8:1:1	6:2:2	1:1:1	8:1:1	6:2:2
Credit	Individual	0.821	0.851	0.846	0.00329	0.00363	0.00308	0.974	0.976	0.976
	TFL [23]	0.847	0.846	0.843	0.00303	0.00294	0.00299	**0.979**	**0.978**	0.978
	F-GBDT-G [24]	0.844	0.850	0.843	0.00301	0.00290	0.00304	0.978	**0.978**	0.977
	eFL-Boost	**0.850**	**0.856**	**0.848**	**0.00275**	**0.00276**	**0.00275**	**0.979**	**0.978**	**0.980**
Breast	Individual	0.962	**0.973**	0.972	0.129	0.101	0.103	0.990	**0.992**	0.991
	TFL [23]	0.971	0.971	0.970	0.110	0.100	0.104	**0.992**	**0.992**	0.992
	F-GBDT-G [24]	0.970	0.968	0.970	0.106	0.113	0.109	**0.992**	0.991	0.992
	eFL-Boost	**0.973**	0.972	**0.975**	**0.101**	**0.0972**	**0.0923**	**0.992**	**0.992**	**0.993**
Biodeg	Individual	0.730	0.759	0.752	0.444	0.431	0.439	0.887	0.905	0.902
	TFL [23]	0.759	0.764	0.774	0.469	0.453	0.458	0.900	0.901	0.901
	F-GBDT-G [24]	0.771	0.761	0.769	0.438	0.436	0.473	0.904	0.903	0.901
	eFL-Boost	**0.776**	**0.772**	**0.776**	**0.393**	**0.405**	**0.397**	**0.907**	**0.906**	**0.908**
German	Individual	0.822	0.826	0.824	0.590*	0.549	0.579	0.760	0.775	0.769
	TFL [23]	0.833	0.827	0.832	0.520	0.615*	0.614	0.786	0.769	0.772
	F-GBDT-G [24]	0.837	0.824	0.831	0.515	0.607*	0.608	0.786	0.763	0.771
	eFL-Boost	**0.841**	**0.836**	**0.835**	**0.508**	**0.526**	**0.522**	**0.788**	**0.785**	**0.784**
Magic	Individual	0.903**	0.908	0.907	0.320*	0.300	0.306	0.924*	0.932	0.930
	TFL [23]	0.906	0.903*	0.906	0.307	0.322**	0.311	0.929	0.925**	0.928
	F-GBDT-G [24]	0.906	0.907	0.907	0.305	0.307	0.308	0.930	0.929	0.930
	eFL-Boost	**0.908**	**0.909**	**0.908**	**0.303**	**0.299**	**0.300**	**0.931**	**0.933**	**0.932**

GBDT is performed for the dataset provided by all data owners without considering privacy preservation. **Individual** indicates the case where training is performed independently by only individual owners. **TFL** denotes the scheme proposed by Zhao *et al.* [23]. **F-GBDT-G** denotes the scheme proposed by Yamamoto *et al.* [24]. **FederBoost** denotes the scheme proposed by Tian *et al.* [26]. Since the learning in **FederBoost** is conducted based on the same procedures in GBDT, it is expected to have the same accuracy as that in **Non-pp**; thus, the results of **Non-pp** are not shown. Moreover, to study the robust property from different aspects, the following experiments are conducted.

B. INFLUENCE OF PARTICIPATING DATA OWNERS ON PERFORMANCE

The most fundamental benefit of FL is that cooperation of multiple data owners improves the prediction model's accuracy. In this experiment, we studied the effectiveness of FL schemes based on the influence of the number of participating owners on the accuracy of the global model. Here, the amount of data for each owner was fixed, the number of participating owners was varied, and the prediction performance was evaluated for each case.

The results are presented in Table 6. The fractions shown in the table indicate the proportion of participating data owners. The number of participating owners was varied while the amount of data for each owner was fixed at 10% of the original dataset. In all schemes, results showed that performance improved as the number of participating owners increased. In particular, eFL-Boost outperforms **Individual**, **TFL**, and **F-GBDT-G** in all cases, and approached **FederBoost** in some datasets. The results show that eFL-Boost has a higher prediction performance than schemes that require less communication costs, and robustness against the number of participating owners.

C. INFLUENCE OF DATA SIZE ON PERFORMANCE

In general, the performance of ML models is affected by the number of training data. Therefore, the FL models to be compared here are evaluated when the numbers of training data in each owner are changed. For this purpose, let us see the performance such that the number of participating owners is fixed.

Table 7 shows the results of the performance comparisons. The fractions in Table 7 have the same meaning as those in Table 6. In this experiment, the number of participating owners is set to three, and the data ratios are set to 1/10, 1/5, and 1/3 against the original dataset. In all GBDT schemes, the prediction performance improved as the amount of data is increased. In particular, eFL-Boost outperforms **Individual**, **TFL** and **F-GBDT-G**.

D. INFLUENCE OF BIAS ON AMOUNT OF DATA OWNED ON PERFORMANCE

In reality, the number of data provided by owners is often different; thus, the FL algorithm should have robust property against the biased data distributions. In this experiment, a strong bias is intentionally imposed on the amount of data, and the robustness against bias is evaluated. Here, we assume that a single owner possesses a large amount of data, and only few data are assigned to the other owners.

The results are shown in Table 8 where the ratios indicate the amount of data owned by each owner to the total amount of data. As seen in Table 8, one owner possesses 80% or 60%

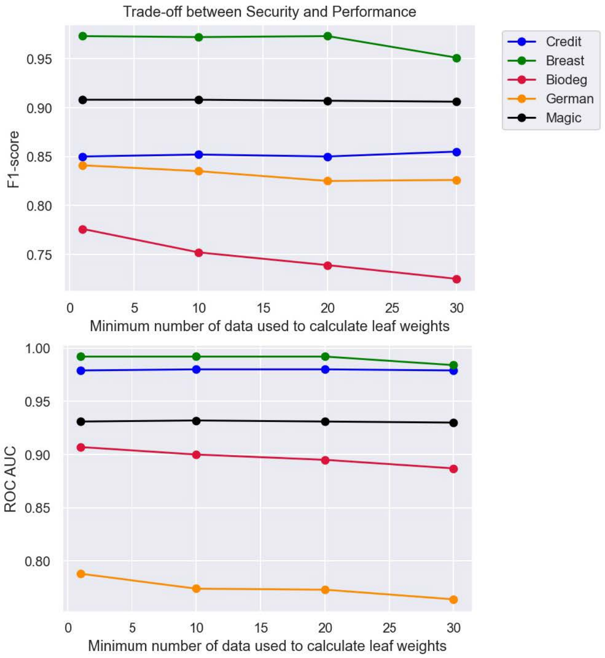

FIGURE 2. Experimental results that verified the influence of restricting the data corresponding to all leaves beyond a certain level on the performance. A negative effect was observed for small data sets, but not for relatively large data sets.

of data and the other owners have only 10% or 20%. Looking at the case of 6:2:2, **TFL** and **F-GBDT-G** tend to have lower performance to **Individual**, while eFL-Boost always attains the best performance. Similar results are obtained for the case of 8:1:1. Hence, unlike **TFL** and **F-GBDT-G**, eFL-Boost always achieves the performance enhancement through FL, even when a strong bias in the amount of data among owners exists.

E. TRADE-OFF BETWEEN SECURITY AND PERFORMANCE

In this experiment, let us discuss the trade-off between privacy preservation and the prediction performance in eFL-Boost. For this purpose, we incur a constraint on the minimum amount of data used in the calculation of leaf weights, which gives a criterion on the security level. By changing the minimum number of data calculated in leaf nodes, the influence to the prediction performance is investigated.

The results are shown in Fig. 2. The horizontal axis indicates the minimum amount of data used to calculate the leaf weights; the larger the value on the horizontal axis, the stronger is the degree of privacy protection. For the F1-score and ROC-AUC, there was no influence on prediction performance for large data sets, i.e., ‘‘Credit’’ and ‘‘Magic;’’ however, some accuracy loss was observed on the small data sets, i.e., ‘‘Breast,’’ ‘‘German,’’ and ‘‘Biodeg’’ data sets. Hence, if the data set is sufficiently large, privacy can be preserved without sacrificing the predictive performance of the model.

VI. CONCLUSION

In this paper, we proposed an FL scheme for GBDT, eFL-Boost, which reduces the communication costs, accuracy loss, and information leakage to other owners. In eFL-Boost, decision tree construction is divided into two stages: tree structure determination and leaf weight calculation. The tree structure is determined by a specific owner, while the leaf weight calculation is cooperatively performed by all owners. The basic ideas in the proposed eFL-Boost were come up with the following facts. The leaf weights are directly related to the GBDT outputs affecting the prediction accuracy; thus, they should be trained with global information. Next, the leaf weight calculation by all owners requires less communications compared to the tree structure determination. Lastly, the low performance caused by local computation within a tree can be compensated through the boosting mechanism. To suppress the privacy leakage from the global model, the lower bound to the minimum number of data trained in the leaf weight calculation should be properly set.

In the experiments, we evaluated eFL-Boost from four perspectives: communication costs, computational complexity, information leakage, and prediction performance. In communication costs, eFL-Boost achieved the same performance as the most efficient scheme, TFL. On the computational complexity, eFL-Boost is comparable to other FL schemes and the conventional GBDT. As for the information leakage, eFL-Boost leaks minimum information to other data owners, and an aggregator only receives low-risk information. The result on the information leakage is almost equivalent to that of TFL, and less than that of FederBoost. In terms of prediction performance, we evaluated the performance of the three schemes for several public data sets, and the influence of the number of participating owners and the amount of data for each data owner on the accuracy was investigated. In almost all cases, eFL-Boost outperforms TFL and F-GBDT-G and has comparable performance to FederBoost.

The following issues still remain in our future work. In some cases, eFL-Boost has lower performance against FederBoost. The main reason for this would be originated from the non-optimality in the tree structure determined by Builder. To solve this, a proper criterion to choose Builder from the data owners is necessary. Second, to reveal stable performance in practical environments, the robustness against strong bias in data distributions should be proved for eFL-Boost. Third, we should conduct more rigorous security analysis so that eFL-Boost can be securely applied to practical sensitive data analysis. The introduction of differential privacy [22], [35] might be one of the solution for this issue. Finally, the concept of eFL-Boost might be possible to extend to the vertical FL.

REFERENCES

[1] J. H. Friedman, ‘‘Greedy function approximation: A gradient boosting machine,’’ *Ann. Statist.*, vol. 29, no. 5, pp. 1189–1232, Oct. 2001.

[2] D. Landgrebe, "A survey of decision tree classifier methodology," *IEEE Trans. Syst., Man Cybern.*, vol. 21, no. 3, pp. 660–674, May 1991.

[3] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2016, pp. 785–794.

[4] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, "LightGBM: A highly efficient gradient boosting decision tree," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 3146–3154.

[5] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, "CatBoost: Unbiased boosting with categorical features," 2017, *arXiv:1706.09516*.

[6] R. Sun, G. Wang, W. Zhang, L.-T. Hsu, and W. Y. Ochieng, "A gradient boosting decision tree based GPS signal reception classification algorithm," *Appl. Soft Comput.*, vol. 86, Jan. 2020, Art. no. 105942.

[7] J. Cheng, G. Li, and X. Chen, "Research on travel time prediction model of freeway based on gradient boosting decision tree," *IEEE Access*, vol. 7, pp. 7466–7480, 2018.

[8] J. Hu and J. Min, "Automated detection of driver fatigue based on EEG signals using gradient boosting decision tree model," *Cognit. Neurodyn.*, vol. 12, no. 4, pp. 431–440, 2018.

[9] J. P. Albrecht, "How the GDPR will change the world," *Eur. Data Protection Law Rev.*, vol. 2, no. 3, pp. 287–289, 2016.

[10] V. S. Verykios, E. Bertino, I. N. Fovino, L. P. Provenza, Y. Saygin, and Y. Theodoridis, "State-of-the-art in privacy preserving data mining," *ACM SIGMOD Rec.*, vol. 33, no. 1, pp. 50–57, 2004.

[11] J. Domingo-Ferrer, "Personal big data, gdpr and anonymization," in *Proc. Int. Conf. Flexible Query Answering Syst.* Cham, Switzerland: Springer, 2019, pp. 7–10.

[12] J. Konečný, H. B. McMahan, F. X. Yu, P. Richtárik, A. T. Suresh, and D. Bacon, "Federated learning: Strategies for improving communication efficiency," 2016, *arXiv:1610.05492*.

[13] T. Yang, G. Andrew, H. Eichner, H. Sun, W. Li, N. Kong, D. Ramage, and F. Beaufays, "Applied federated learning: Improving Google keyboard query suggestions," 2018, *arXiv:1812.02903*.

[14] L. Li, Y. Fan, M. Tse, and K.-Y. Lin, "A review of applications in federated learning," *Comput. Ind. Eng.*, vol. 149, Nov. 2020, Art. no. 106854.

[15] Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated machine learning: Concept and applications," *ACM Trans. Intell. Syst. Technol.*, vol. 10, no. 2, pp. 1–19, 2019.

[16] Y. Liu, Y. Liu, Z. Liu, Y. Liang, C. Meng, J. Zhang, and Y. Zheng, "Federated forest," *IEEE Trans. Big Data*, early access, May 7, 2020, doi: 10.1109/TBDATA.2020.2992755.

[17] R.-H. Hsu, Y.-C. Wang, C.-I. Fan, B. Sun, T. Ban, T. Takahashi, T.-W. Wu, and S.-W. Kao, "A privacy-preserving federated learning system for Android malware detection based on edge computing," in *Proc. 15th Asia Joint Conf. Inf. Secur. (AsiaJCIS)*, Aug. 2020, pp. 128–136.

[18] L. T. Phong, Y. Aono, T. Hayashi, L. Wang, and S. Moriai, "Privacy-preserving deep learning via additively homomorphic encryption," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 5, pp. 1333–1345, May 2018.

[19] H. Kim, J. Park, M. Bennis, and S.-L. Kim, "Blockchained on-device federated learning," *IEEE Commun. Lett.*, vol. 24, no. 6, pp. 1279–1283, Jun. 2020.

[20] K. Cheng, T. Fan, Y. Jin, Y. Liu, T. Chen, D. Papadopoulos, and Q. Yang, "SecureBoost: A lossless federated learning framework," 2019, *arXiv:1901.08755*.

[21] Y. Liu, Z. Ma, X. Liu, S. Ma, S. Nepal, and R. Deng, "Boosting privately: Privacy-preserving federated extreme boosting for mobile crowdsensing," 2019, *arXiv:1907.10218*.

[22] Q. Li, Z. Wu, Z. Wen, and B. He, "Privacy-preserving gradient boosting decision trees," in *Proc. AAAI Conf. Artif. Intell.*, vol. 34, no. 1, 2020, pp. 784–791.

[23] L. Zhao, L. Ni, S. Hu, Y. Chen, P. Zhou, F. Xiao, and L. Wu, "InPrivate digging: Enabling tree-based distributed data mining with differential privacy," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, Apr. 2018, pp. 2087–2095.

[24] F. Yamamoto, L. Wang, and S. Ozawa, "New approaches to federated Xgboost learning for privacy-preserving data analysis," in *Proc. Int. Conf. Neural Inf. Process.* Cham, Switzerland: Springer, 2020, pp. 558–569.

[25] Q. Li, Z. Wen, and B. He, "Practical federated gradient boosting decision trees," in *Proc. AAAI Conf. Artif. Intell.*, vol. 34, no. 4, 2020, pp. 4642–4649.

[26] Z. Tian, R. Zhang, X. Hou, J. Liu, and K. Ren, "FederBoost: Private federated learning for GBDT," 2020, *arXiv:2011.02796*.

[27] P. Paillier, "Public-key cryptosystems based on composite degree residuosity classes," in *Advances in Cryptology* (Lecture Notes in Computer Science), vol. 1592. Springer, 1999, pp. 223–238.

[28] C. Gentry, "Fully homomorphic encryption using ideal lattices," in *Proc. 41st Annu. ACM Symp. Symp. Theory Comput. (STOC)*, 2009, pp. 169–178.

[29] Kaggle. *Credit Card Fraud Detection*. [Online]. Available: https://www.kaggle.com/mlg-ulb/creditcardfraud

[30] UCI. *Breast Cancer Wisconsin Data Set*. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/breast+cancer+wisconsin+(diagnostic)

[31] UCI. *QSAR Biodegradation Data Set*. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/QSAR+biodegradation

[32] UCI. *German Credit Data*. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/statlog+(german+credit+data)

[33] UCI. *MAGIC Gamma Telescope Data Set*. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/magic+gamma+telescope

[34] B. L. Welch, "The generalization of 'student's' problem when several different population varlances are involved," *Biometrika*, vol. 34, nos. 1–2, pp. 28–35, 1947.

[35] N. Li, W. Qardaji, and D. Su, "On sampling, anonymization, and differential privacy or, k-anonymization meets differential privacy," in *Proc. 7th ACM Symp. Inf., Comput. Commun. Secur. (ASIACCS)*, 2012, pp. 32–33.

FUKI YAMAMOTO received the B.E. degree in engineering from Kobe University, Japan, in 2020, where he is currently pursuing the M.E. degree in engineering with the Graduate School. His research interests include machine learning and its applications on privacy-preserving data mining.

SEIICHI OZAWA (Senior Member, IEEE) received the Dr.Eng. degree in computer science from Kobe University, Japan. He is currently the Director of the Center for Mathematical and Data Sciences along with a Full Professor with the Department of Electrical and Electronic Engineering, Graduate School of Engineering, and the Center for Advanced Medical Engineering Research & Development, Kobe University. He has published more than 160 journals and conference papers, and book chapters/monographs. His current research interests include machine learning, incremental learning, big data analytics, cybersecurity, text mining, computer vision, and privacy-preserving machine learning. He is currently an Associate Editor of IEEE TRANSACTION ON NEURAL NETWORKS AND LEARNING SYSTEMS, IEEE TRANSACTION ON CYBERNETICS, and two international journals. He is the Vice-President of Membership of International Neural Network Society, the President of Asia Pacific Neural Network Society, and the Vice-President of Japan Neural Network Society. He is a member of Neural Networks TC of IEEE CI Society.

LIHUA WANG received the B.S. degree in mathematics from Northeast Normal University, China, in 1988, the M.S. degree in mathematics from the Harbin Institute of Technology, China, in 1994, and the Ph.D. degree in engineering from the University of Tsukuba, Japan, in 2006. She is currently a Senior Researcher with the Cybersecurity Research Institute, National Institute of Information and Communications Technology (NICT), Japan. Her research interests include cryptography and its applications on privacy-preserving machine learning.