Explainable Semantic Federated Learning Enabled Industrial Edge Network for Fire Surveillance

Li Dong, Yubo Peng, Feibo Jiang, Senior Member, IEEE, Kezhi Wang, Senior Member, IEEE, and Kun Yang, Fellow, IEEE

Abstract-In fire surveillance, Industrial Internet of Things (IIoT) devices require transmitting large monitoring data frequently, which leads to huge consumption of spectrum resources. Hence, we propose an Industrial Edge Semantic Network (IESN) to allow HoT devices to send warnings through Semantic communication (SC). Thus, we should consider (1) Data privacy and security. (2) SC model adaptation for heterogeneous devices. (3) Explainability of semantics. Therefore, first, we present an eXplainable Semantic Federated Learning (XSFL) to train the SC model, thus ensuring data privacy and security. Then, we present an Adaptive Client Training (ACT) strategy to provide a specific SC model for each device according to its Fisher information matrix, thus overcoming the heterogeneity. Next, an Explainable SC (ESC) mechanism is designed, which introduces a leakyReLUbased activation mapping to explain the relationship between the extracted semantics and monitoring data. Finally, simulation results demonstrate the effectiveness of XSFL.

Index Terms—Fire surveillance; federated learning; semantic communication; explainable AI

I. INTRODUCTION

Every year, industrial fires pose a serious threat to human lives and property due to human errors or system failures. The risk of fire is further heightened by high infrastructure concentration and stored inventory [1]. When a fire occurs, it is crucial to promptly issue warnings and initiate rescue missions to minimize property damage and increase the chances of saving lives. However, traditional manual monitoring methods are inconvenient, scattered, and come with certain drawbacks such as high costs and poor network connectivity [2]. Therefore, effectively and timely sending critical information upon detecting a fire presents a significant challenge for traditional manual monitoring methods.

Yubo Peng (pengyubo@hunnu.edu.cn) is with the School of Intelligent Software and Engineering, Nanjing University, Suzhou, China.

Feibo Jiang (jiangfb@hunnu.edu.cn) is with the Hunan Provincial Key Laboratory of Intelligent Computing and Language Information Processing, Hunan Normal University, Changsha 410081, China.

Kezhi Wang (Kezhi.Wang@brunel.ac.uk) is with the Department of Computer Science, Brunel University London, UK.

Kun Yang (kyang@ieee.org) is with the School of Intelligent Software and Engineering, Nanjing University, Suzhou, China.

With the development of Artificial Intelligence (AI) and the Industrial Internet of Things (IIoT), AI-empowered IIoT devices play an important role in catering to a multitude of problems (e.g., transportation, healthcare, and disaster management, etc.) efficiently and reliably [3]. Simultaneously, it is also used for fire surveillance. In [4], the authors proposed an efficient system, which was based on a lightweight Deep Learning (DL) model, for fire detection in uncertain surveillance scenarios. In [5], the authors introduced a vision-based fire detection framework that focuses on monitoring private spaces while prioritizing occupant privacy. An innovative solution for continuous forest surveillance and fire detection was developed in [6] using a drone-based system integrated with AI capabilities, specifically designed for forest firefighting purposes. In [7], a novel approach leveraging IIoT technology was proposed to detect fires and promptly notify relevant parties. This method ensured accurate solutions for fire surveillance systems.

1

However, IIoT devices require frequent transmissions of large monitoring data when the conflagration occurs, leading to a huge consumption of spectrum resources. Semantic Communication (SC) is viewed as one of the intelligent solutions to physical layer communication. Different from traditional ways, SC transmits only semantic features required for the respective task [8]. The semantic feature is generally defined as the symbols with much fewer dimensions extracted from the raw data. For example, in fire surveillance, the IIoT devices only extract the semantic information from the surveillance image with the deployed SC encoder, then transmit it to the Edge Server (ES) deployed on the Base Station (BS). The ES decodes the received semantic information with the deployed SC decoder and obtains the prediction results. Since the transferred data is greatly reduced in SC, the consumption of spectrum resources and communication delay are also greatly decreased. The high-accuracy SC model based on DL is critical for the performance of SC, while several challenges should be considered as follows:

 Data privacy and security for industrial applications: Most data in IIoT devices are sensitive for referring personal information and business secrets [9]. Hence, privacy security has become a critical issue. However, all data is required to be stored centrally for fire surveillance, and the detection model is trained and carried out in the traditional centralization method, which brings overwhelming communication costs and huge risks of data leakage for edge users.

This work was supported in part by the National Natural Science Foundation of China under Grant 41904127 and 62132004, in part by the Hunan Provincial Natural Science Foundation of China under Grant 2024JJ5270, in part by the Open Project of Xiangjiang Laboratory under Grant 22XJ03011, and in part by the Scientific Research Fund of Hunan Provincial Education Department under Grant 22B0663. (Corresponding authors: Yubo Peng, Feibo Jiang.)

Li Dong (Dlj2017@hunnu.edu.cn) is with the School of Computer Science, Hunan University of Technology and Business, Changsha 410205, China, and also with the Xiangjiang Laboratory, Changsha 410205, China.

- 2) High training delay for heterogeneous devices: In the edge network, the IIoT devices are heterogeneous. For well-resourced devices, a sophisticated global model could ensure the model's accuracy. However, such a global model may bring a high training delay for the limited-resourced devices, which is not accepted for fire surveillance [10]. Although a compact global model could ensure the training delay for all participating IIoT devices, the model accuracy is difficult to guarantee, especially for well-resourced IIoT devices [11].
- 3) Explainability of extracted fire semantics for succors: The erroneous predictions by DL models may potentially result in fatal accidents in recent years. Hence, succors couldn't trust DL when making life-critical decisions for it is nontransparent and unexplainable [12]. Therefore, the DL-based extracted semantics, representing the features of fire, must be transparent and explainable.

To address the above issues, we propose an Industrial Edge Semantic Network (IESN) that reduces the transmitted information and thus decreases the communication overheads of IIoT devices. Our main contributions can be summarized as follows:

- We introduce the architecture of IESN, in which the SC is used as the communication way between IIoT devices and ES. SC can make devices transmit only the extracted semantic information rather than the raw data. Thus, the IESN can reduce the transmitted information and improve the communication efficiency in edge networks. Furthermore, we propose an eXplainable Semantic Federated Learning (XSFL) for IIoT devices and ES to co-train the SC model, and this decentralized learning method at the edge can protect the privacy and security of industrial data.
- 2) We present the Adaptive Clients Training (ACT) strategy in the ES. Specifically, we first cluster the IIoT devices according to their local resources. Then, for each IIoT device, ES publishes a specific FL model with a proportion of frozen parameters according to the Fisher information of its local weights. The fewer local resources the IIoT device owns, the more the frozen ratio of the parameters, and thus the FL training delay is decreased. Therefore, the local FL models can better adapt to heterogeneous devices with different resources, and the second challenge is solved.
- 3) We propose the Explainable SC (ESC) mechanism in edge monitor devices to provide explainability of the extracted semantics based on DL. Concretely, the ESC uses semantic-specific gradient information to localize important semantic features. Then, we introduce a leakyReLU-based activation mapping to combine the localized features with the raw monitoring data to develop a high-resolution heatmap visualization. Thus, the decision-makers can understand the relative between the monitoring data and extracted semantics. Therefore, we address the third challenge.

The rest of this paper is organized as follows. The system

model is introduced in Section II. The proposed XSFL is described in Section III. The complexity analyses are introduced in Section IV. Section V presents the numerical results. The work summary and future works are described in Section VI.

II. SYSTEM MODEL

Fig. 1 (a) shows the architecture of the proposed IESN in fire surveillance. For a spectrum resource-limited wireless network, we deploy a distributed SC system which consists of N IIoT devices, and one ES deployed on the BS. The set of IIoT devices is denoted as $\mathcal{N} = \{1, 2, ..., N\}$. The IIoT devices only need to transmit the semantic information to the BS, rather than the raw monitoring data, then the semantic information is decoded in the ES. Once the decoding results of the received semantics represent the occurrence of a fire, the decision-makers should take the corresponding actions, e.g., dispatching ambulances and fire trucks. To enable IIoT devices can extract the semantic information, the SC encoder is deployed on each IIoT device, while the SC decoder is deployed on the ES to decode the received semantic information. The architecture of the SC model is introduced in Section III-A. In addition, we should consider the impairments of the physical channel between IIoT devices and the BS.



Fig. 1: The illustration of IESN for fire surveillance. (a) Distributed SC system over the wireless network. (b) The architecture of the SC model.

A. Semantic Communication Model in IESN

The encoding and decoding of the semantic information are critical for the accuracy of SC, hence we must consider the architecture design of the SC model for IESN. As shown in Fig. 1 (b), the SC model mainly consists of three parts: SC encoder, physical channel, and SC decoder. The SC encoder

includes the semantic and channel encoder, and it is deployed on each IIoT device. The SC decoder consists of the channel and semantic decoder, and it is deployed on ES. Note the above parts of the SC model are built based on deep neural networks, and different network layers (e.g. convolutional layer, transformer layer, etc.) can be used according to different semantic tasks [13].

The SC encoder encodes the input raw data x and extracts the semantic information. The encoding result can be expressed as:

$$\mathbf{X} = C\left(S\left(x, \boldsymbol{\vartheta}\right), \boldsymbol{\alpha}\right), \tag{1}$$

where **X** represents the extracted semantic information; $S(\cdot)$ represents the semantic encoder with model parameters ϑ and $C(\cdot)$ is the channel encoder with model parameters α .

When transmitted over the physical channel, \mathbf{X} suffers transmission impairments that include distortion and noise. The transmission process can be given by:

$$\mathbf{Y} = \mathbf{h}\mathbf{X} + \mathbf{N},\tag{2}$$

where Y represents the received semantic information; h represents the channel gain between the transmitter and the receiver; N is Additive White Gaussian Noise (AWGN). For end-to-end training of the encoder and decoder, the physical channel must allow backpropagation. Therefore, the physical channel can be simulated by neural networks [14], [15].

The SC decoder performs the decoding of the received semantic information \mathbf{Y} and alleviates transmission impairments. The decoding result \hat{x} can be given by:

$$= S^{-1} \left(C^{-1}(\mathbf{Y}, \boldsymbol{\beta}), \boldsymbol{\delta} \right),$$

where $C^{-1}(\cdot)$ represents the channel decoder with model parameters β ; $S^{-1}(\cdot)$ is the semantic decoder with model parameters δ .

In this paper, we approach the SC task as image classification. As such, we utilize the cross-entropy (CE) loss function in the SC system:

$$\mathcal{L}_{\rm CE}(\mathbf{y}, \hat{\mathbf{y}}) = -\sum_{i=1}^{M} y_i \log\left(\hat{y}_i\right),\tag{4}$$

where $\mathbf{y} = [y_1, y_2, ..., y_M]$ represents the array of labels, if the input data x belongs to the *i*-th class, then $y_i = 1$ otherwise 0; $\hat{\mathbf{y}} = [\hat{y}_1, \hat{y}_2, ..., \hat{y}_M]$ represents the array of the probabilities of all classes, \hat{y}_i represents the probability that x is predicted as the *i*-th class; M is the total number of categories. Moreover, classification-related metrics, such as accuracy, precision, etc., can be used to evaluate the quality of SC [16].

B. Federated Learning Model in IESN

 \hat{r}

We denote the local dataset of the IIoT device $n \ (n \in \mathcal{N})$ as $\mathcal{D}_n = \{(x_{n,1}, y_{n,1}), (x_{n,2}, y_{n,2}), ..., (x_{n,D_n}, y_{n,D_n})\}$, where D_n is the number of collected samples in \mathcal{D}_n , $x_{n,i}$ is the *i*-th sample and $y_{n,i}$ is the corresponding label.

In the *t*-th communication round of FL, the local loss function of the IIoT device $n_{\rm D}$ can be calculated as:

$$F_{n}(\mathbf{w}_{n,t}) = \frac{1}{D_{n}} \sum_{i=1}^{D_{n}} f(\mathbf{w}_{n,t}, x_{n,i}, y_{n,i}),$$
(5)

where $f(\mathbf{w}_{n,t}, x_{n,i}, y_{n,i})$ is the loss function for the *i*-th sample $(x_{n,i}, y_{n,i})$ in \mathcal{D}_n ; $\mathbf{w}_{n,t}$ represents the weight assigned to the local FL model of the IIoT device n in the *t*-th

communication round. Moreover, $\mathbf{w}_{n,t}$ includes all the SC parameters, namely $\mathbf{w}_{n,t} = (\boldsymbol{\alpha}_{n,t}, \vartheta_{n,t}, \beta_{n,t}, \delta_{n,t})$, where $\boldsymbol{\alpha}_{n,t}, \vartheta_{n,t}, \beta_{n,t}, \delta_{n,t}$ represent the parameters of the SC model of the IIoT device *n* from the channel encoder to the semantic decoder in the *t*-th communication round.

To ensure data privacy and security, the global FL model is updated in the ES by aggregating the weights $\mathbf{w}_{n,t}$ of the local FL models from various IIoT devices. The update of the global FL model can be calculated by combining and processing these weights as follows:

$$\mathbf{w}_{g,t} = \frac{1}{D} \sum_{n=1}^{N} D_n \mathbf{w}_{n,t},\tag{6}$$

where $\mathbf{w}_{g,t}$ is the updated weight of the global FL model; $D = \sum_{n=1}^{N} D_n$ represents the sum of local data. Note that $\mathbf{w}_{g,t}$ shares the same architecture as $\mathbf{w}_{n,t}$.

FL aims to minimize the local FL loss of all devices with an optimal global model $\mathbf{w}_{g,t}$, thus achieving global optimization. Hence, the global loss function of FL can be given by:

$$F_g\left(\mathbf{w}_{g,t}\right) = \frac{1}{N} \sum_{n=1}^{N} F_n\left(\mathbf{w}_{g,t}\right),\tag{7}$$

where $F_n(\mathbf{w}_{g,t})$ represents that the local FL loss based on $\mathbf{w}_{g,t}$. In this paper, the minimization of $F_g(\mathbf{w}_{g,t})$ is the training goal of the SC model in IESN.

C. Transmission Model in IESN

(3)

In IESN, the uplink data rate of the IIoT device n via the channel j can be calculated as [17]:

$$R_{n,j}^{\rm U}(t) = B^{\rm U} \log_2 \left\{ 1 + \frac{P^{\rm U} h_{n,j}(t)}{I_{n,j}^{\rm U}(t) + \sigma} \right\},\tag{8}$$

where $P^{\rm U}$ denotes the transmission power of the IIoT device; $h_{n,j}(t)$ represents the channel gain when device *n* communicates with the ES via channel *j*; $I_{n,j}^{\rm U}(t)$ refers to the co-channel interference experienced; σ indicates the noise power spectral density; and $B^{\rm U}$ signifies the bandwidth utilized during the uploading process.

Similarly, when the global model is transmitted to the device n by the ES, the data rate of the downlink can be expressed as:

$$R_{n,j}^{\rm D}(t) = B^{\rm D} \log_2 \left\{ 1 + \frac{P^{\rm D} h_{n,j}(t)}{I_{n,j}^{\rm D}(t) + \sigma} \right\},\tag{9}$$

where $P^{\rm D}$ represents the transmission power of the ES; $B^{\rm D}$ signifies the bandwidth of the downlink channel; and $I_{n,j}^{\rm D}(t)$ denotes the interference experienced during the downloading process.

Hence, the uplink and downlink transmission delays can be respectively derived:

$$d_{n,j}^{\rm U}(t) = \frac{Z(\mathbf{w}_{n,t})}{R_{n,j}^{\rm U}(t)}, \quad d_{n,j}^{\rm D}(t) = \frac{Z(\mathbf{w}_{n,t})}{R_{n,j}^{\rm D}(t)}, \tag{10}$$

where $Z(\mathbf{w}_{n,t})$ denotes required bits for transmitting $\mathbf{w}_{n,t}$ over wireless links.

D. Computation Model in IESN

We presume that each IIoT device is furnished with a CPU to perform local training tasks. The computational capacity

of each device is assessed by its CPU frequency, denoted as $f_{n,t}$. In addition, $\Phi_n(\mathbf{w}_{n,t})$ represents the number of CPU cycles required to carry out the forward-backward propagation process utilizing a single data point for the local model $\mathbf{w}_{n,t}$. Consequently, for the *t*-th communication round, the calculation delay of the local FL model can be derived [17]:

$$d_n^{\mathrm{L}}(t) = \frac{D_n \Phi_n(\mathbf{w}_{n,t})G}{f_{n,t}},\tag{11}$$

where G is the local training iterations.

E. Delay Analysis and Optimization Objective

Assuming the ES possesses ample computational resources, the delay caused by global model aggregation is considered insignificant [17]. Therefore, the FL training delay for the n-th device during the t-th communication round can be expressed as follows:

$$d_{n,t} = d_{n,j}^{\rm U}(t) + d_{n,j}^{\rm D}(t) + d_n^{\rm L}(t).$$
(12)

Since the global aggregation step is synchronous in FL [18], during the t-th communication round, the total delay of the FL training can be expressed as:

$$d_{t} = \max_{n \in \mathcal{N}} \{ \min \{ d_{n,t}, d_{\max} \} \},$$
(13)

 d_{max} signifies the maximum permitted delay for each IIoT device *n* when executing the FL algorithm. Specifically, if $d_{n,t} > d_{\text{max}}$, IIoT device *n* cannot participate in the *t*-th round of FL training. The primary objective of this paper is to minimize both the global FL loss and the FL training delay.

III. THE XSFL FRAMEWORK FOR IESN

To achieve the IESN for fire surveillance, a high-accuracy SC model is critical. Hence, the XSFL framework is proposed, in which we apply the ACT strategy and ESC mechanism in the ES and IIoT devices, respectively.

A. XSFL Framework

In this subsection, we introduce the details of the XSFL framework. We assume that XSFL starts from a certain round t, as shown in Fig. 2, the workflow of the XSFL is as follows:

1) Parameters Publication: To make the local FL models adaptive to the heterogeneous IIoT devices, we perform the ACT strategy on the global FL model $\mathbf{w}_{g,t}$ in ES and the global parameters that require training for each IIoT device are published. Afterward, according to the published parameters, the global FL models are processed specifically, and they are broadcasted to the corresponding IIoT devices for updating and starting the t + 1-th FL:

$$\mathbf{w}_{n,t+1} = \hat{\mathbf{w}}_{g,t+1}^n,\tag{14}$$

where $\mathbf{w}_{n,t+1}$ denotes the local FL model during the (t+1)-th communication round, while $\hat{\mathbf{w}}_{g,t+1}^n$ represents the global FL model specific to IIoT device n. It is important to note that both $\hat{\mathbf{w}}_{g,t+1}^n$ and $\mathbf{w}_{n,t+1}$ share the same architectural design.

2) Local Training and Semantic Explanation: Each IIoT device n performs local training with the local FL model $\mathbf{w}_{n,t+1}$. After training, each IIoT device applies the ESC mechanism to explain the fire semantics from the local FL model in a visual way, which reveals the mapping between semantic features extracted by the SC encoder and input monitoring data.

3) Parameters Aggregation: When the trained local FL models from all IIoT devices are uploaded to the ES, the ES performs parameters aggregating. Afterward, we obtain the aggregated global FL model. Assuming the total number of communication rounds is T, the local training epochs of each IIoT device is G, and the proposed XSFL framework is summarized in Algorithm 1.

Algorithm 1 XSFL framework

Input: T, G.

Output: $\mathbf{w}_{g,T}$.

- 1: for each communication round $t \in T$ do
- 2: Edge server do
- 3: Aggregate local FL models from IIoT devices according to Eq. (6).
- 4: Perform ACT strategy on the global FL model according to Algorithm 2.
- 5: Broadcast the specific global FL models $\hat{\mathbf{w}}_{g,t+1}^n$ corresponding to each IIoT device for updating according to Eq. (14).
- 6: Edge device do
- 7: for each local epoch in G do
- 8: Update $\mathbf{w}_{n,t}$ by minimizing Eq. (5).
- 9: end for
- 10: Explain the extracted semantics with the ESC mechanism according to **Algorithm 3**.

11: end for

B. Adaptive Clients Training Strategy

Practically, IIoT devices are usually heterogeneous, which makes different IIoT devices with disparate local resources have different FL training delays. From Eq. (10) and Eq. (11), it becomes evident that the FL training delay is affected by the size of the FL model parameters. Moreover, the model size also plays a significant role in influencing the model's accuracy. Hence, there exists a tradeoff between the model accuracy and FL training delay. Concretely, a sophisticated model can fully utilize data resources for well-resourced IIoT devices, and thus the model accuracy is ameliorative. For limited-resourced IIoT devices, a relatively compact model can meet the requirement for low model accuracy but reduce the FL training delay.

Therefore, we present the ACT strategy to achieve the FL model match to different IIoT devices thus reducing the FL training delay while ensuring model accuracy. Concretely, the IIoT devices just train a proportion of the model parameters, and the proportion is determined by their local data volumes, and the adjusted parameters are determined by the Fisher information matrix. The workflow of the ACT strategy is as follows:

1) Device Classification: We perform clustering of the IIoT devices according to their local data volumes. Specifically, we use the following formulation as the distance evaluation functions [19]:

$$\psi = \sum_{i=1}^{C} \sum_{n \in \mathcal{C}_i} \|D_n, m_i\|_2^2,$$
(15)



Fig. 2: The illustration of the XSFL framework. (a) ACT strategy; (b) ESC mechanism

where C denotes the total number of clusters; C_i is the *i*-th cluster; and m_i represents the medoid of the *i*-th cluster with $m_i \in C_i$. The medoid m_i can be computed by [19]:

$$m_{i} = \arg\min_{n \in \mathcal{C}_{i}} \sum_{j \in \mathcal{C}_{i}} \|D_{n}, D_{j}\|_{2}^{2}.$$
 (16)

Hence, for each IIoT device n, its belonging cluster can be represented as a set $C_i^n (1 \le i \le C)$.

2) Parameters Proportion: The proportion ζ_n of parameters that need to be trained for each IIoT device is also divided into *C* classes according to their belonging clusters C_i^n . For the same cluster of IIoT devices, the average of their data volumes is used to represent the data volume of the whole class, hence ζ_n can be calculated by:

$$\zeta_n = \frac{\text{Mean}(\tilde{\mathcal{C}}_i^n)}{D_{\text{max}}},\tag{17}$$

where $Mean(\mathcal{C}_i^n)$ represents the average value of all elements in \mathcal{C}_i^n ; D_{max} is the max data volumes. Well-resourced IIoT devices are required to train more FL model parameters to maintain model accuracy, while limited-resourced devices focus on training fewer parameters of the FL model to reduce the FL training delay.

3) Weights Freezing: By employing the Fisher information matrix of the local FL model $\mathbf{w}_{n,t}$, we can estimate the importance of each parameter in the global FL model $\mathbf{w}_{g,t}$ relative to the local FL model $\mathbf{w}_{n,t}$. This can be expressed as:

 $\mathbf{I}_{g,t}^{n} = (\mathbf{w}_{g,t} - \mathbf{w}_{n,t})^{\top} \operatorname{diag}(\mathcal{I}_{n}(\mathbf{w}_{n,t}))(\mathbf{w}_{g,t} - \mathbf{w}_{n,t})$ (18) where $\operatorname{diag}(\cdot)$ is the operation of obtaining diagonal elements; $\mathcal{I}_{n}(\mathbf{w}_{n,t})$ represents the Fisher information matrix of the local FL model $\mathbf{w}_{n,t}$. Then, a part of parameters of $\mathbf{I}_{g,t}^{n}$ with the smallest absolute value and their indexes are selected, and each IIoT device *n* gets the corresponding global FL model $\hat{\mathbf{w}}_{g,t}^{n}$ according to the indexes $\mathbf{I}_{g,t,\text{sel}}^{n}$. This can be expressed as follows:

$$\mathbf{I}_{g,t,\text{sel}}^{n} = \text{Select}(\mathbf{w}_{g,t}, 1 - \zeta_{n}), \tag{19}$$

$$\hat{\mathbf{w}}_{g,t}^{n} = \text{Freeze}(\mathbf{w}_{g,t}, \mathbf{I}_{g,t,\text{sel}}^{n}), \qquad (20)$$

where function Select($\mathbf{w}_{g,t}, 1 - \zeta_n$) represents the operation that select $1 - \zeta_n$ proportion weights of the global FL model $\mathbf{w}_{g,t}$ with the smallest absolute values of $\mathbf{I}_{g,t}^n$ and output their indexes $\mathbf{I}_{g,t,\text{sel}}^n$; The function Freeze($\mathbf{w}_{g,t}, \mathbf{I}_{g,t,\text{sel}}^n$) represents that in the global FL model $\mathbf{w}_{g,t}$, the weights of the selected parameters according to $\mathbf{I}_{g,t,\text{sel}}^n$ are frozen. The description of the ACT strategy is summarized in **Algorithm** 2.

4) Hessian Interpretation: Since the Fisher information matrix approximates the Hessian $\mathbf{H}_{n,t}$ of the local loss $F_n(\mathbf{w}_{n,t})$ at $\mathbf{w}_{n,t} = \mathbf{w}_{g,t}$ under some regularity conditions [20]:

$$\mathbf{I}_{g,t}^{n} \approx \frac{1}{2} (\mathbf{w}_{g,t} - \mathbf{w}_{n,t})^{\top} \mathbf{H}_{n,t} (\mathbf{w}_{g,t} - \mathbf{w}_{n,t}), \qquad (21)$$

where $\mathbf{H}_{n,t}$ represents the Hessian matrix of $\mathbf{w}_{n,t}$, and Eq. (21) is equal to the loss function. Therefore, the weights with small values in $\mathbf{I}_{g,t}^n$ can be frozen, and the weights with large values in $\mathbf{I}_{g,t}^n$ are exactly the goals we want to adjust to minimize the loss function.

Algorithm 2 ACT strategy

Input: $\mathbf{w}_{g,t}$.

Output: $\hat{\mathbf{w}}_{g,t}^n$.

- 1: Cluster the IIoT devices according to Eqs. (15) and (16), then each IIoT device *i* obtains the belonging cluster C_i^n .
- 2: for each IIoT device $n \in \mathcal{N}$ do
- 3: Calculate the trained parameters proportion ζ_n according to Eq. (17).
- 4: Obtain the specific global FL model $\hat{\mathbf{w}}_{g,t}^n$ for the IIoT device *n* according to Eqs. (18) (20).

5: end for

In the ACT strategy, the partial weights of the global FL model are frozen to accommodate the local data volumes of the heterogeneous IIoT devices. This approach enables the local FL model to adapt better to the local data, thereby improving its accuracy. Simultaneously, the number of trainable local parameters in the IIoT devices is reduced, leading to a corresponding decrease in the FL training delay [21], [22].

C. Explainable SC Mechanism

The inaccurate fire semantics may lead to incorrect decisions in fire relief. Hence, the explainability of extracted fire semantics is critical for being trustworthy [23]. As a solution, the ESC mechanism is applied in IIoT devices to explain the extracted fire semantics from local FL models (namely the SC models). Specifically, we introduce a novel leakyReLU-based activation mapping to explain the relative between the semantic features extracted by the SC encoder and input monitoring data. The ESC mechanism employs semantic feature-specific gradient information to identify significant regions within input images. By visualizing this information, practitioners can gain insights into the underlying rationale for a particular prediction made by the SC model. As shown in Fig. 2(b), the steps of the ESC in a visual way are illustrated as follows:

1) Semantic Gradient Computation: We assume the semantic features $\mathbf{Q} = [Q_1, Q_2, ..., Q_L]$ is a one-dimensional vector, where Q_l represents weight value of the *l*-th feature. Then, we compute the gradient of the semantic feature Q_l to the feature map activation A^k for the kernel k of the last convolution layer, the calculated formulation is as follows:

$$G_l = \frac{\partial Q_l}{\partial A^k},\tag{22}$$

where G_l represents the gradients for the *l*-th semantic feature. In this step, note that all values calculated depend on the input monitoring data and the weights of the SC model are fixed.

2) Feature Significance Calculation: We compute the neuron importance weights, ω_k^l , by taking a weighted average of the pixel-wise gradients along both height h and width w. The calculation of ω_k^l can be expressed as:

$$\omega_{k}^{l} = \sum_{i=1}^{h} \sum_{j=1}^{w} \varrho_{i,j}^{k,l} G_{l}, \qquad (23)$$

where $\varrho_{i,j}^{k,l}$ is the weighting coefficients for the pixel-wise gradients for semantic feature Q_l and feature map A^k , which can be calculated by:

$$\varrho_{i,j}^{k,l} = \frac{\frac{\partial^2 Q_l}{\left(\partial A_{i,j}^k\right)^2}}{2\frac{\partial^2 Q_l}{\left(\partial A_{i,j}^k\right)^2} + \sum_{a=1}^h \sum_{b=1}^w A_{a,b}^k \left\{\frac{\partial^3 Q_l}{\left(\partial A_{i,j}^k\right)^3}\right\}}$$
(24)

where $A_{i,j}^k$ and $A_{a,b}^k$ both represent someone pixel in A^k . Eq. (24) shows that the weight of $A_{i,j}^k$ is affected by itself and the other pixels $A_{a,b}^k$. Consequently, ω_k^l signifies a partial linearization of the deep network downstream from A^k , and encapsulates the importance of feature map k concerning the *l*-th semantic feature [24].

3) Visual Heatmap Generation: The visual heatmap that ESC generated is a weighted combination of feature maps. We introduce Leaky ReLU to retain the information of the negative values in the semantic features, thus solving the "Dead Neuron" problem [25] and ensuring the ESC mechanism generates remarkable heatmaps for the semantic explanation. The localization map P_l with Leaky ReLU for semantic feature Q_l can be given by:

$$P_{l} = \text{leakyReLU}\left(\sum_{e=1}^{k} \omega_{k}^{l} A^{k}\right) = \max(0, \sum_{e=1}^{k} \omega_{k}^{l} A^{k}) + \alpha * \min(0, \sum_{e=1}^{k} \omega_{k}^{l} A^{k}),$$
(25)

where $\alpha \in (0, 1)$ is the slope for negative input. Finally, we obtain the complete localization map **P** for all semantics **S**, where $\mathbf{P} = \{P_1, P_2, ..., P_L\}$. Then, **P** is normalized and thus

results in a heatmap for visualization. **Algorithm** 3 shows the description of the ESC mechanism.

Algorithm 3	ESC mechanism	
-------------	---------------	--

Input: S.

Output: P.

- 1: for each semantic feature $Q_l \in \mathbf{S}$ do
- 2: Compute the gradient G_l of the semantic feature Q_l according to Eq. (22).
- 3: Obtain the neuron importance weights ω_k^l of the gradient G_l according to Eqs. (23) and (24).
- 4: Calculate P_l according to Eq. (25).

5: **end for**

With the ESC mechanism, we achieve the explainability of the fire semantics from the SC model. The generated heatmap by ESC captures the relationship between the semantic features and input monitoring data, which helps the succors interpret the prediction results and thus improves the trustworthiness of SC. In addition, the ESC mechanism can be used with the SC model locally on IIoT devices and without any pre-training, hence there is no extra computation cost for using ESC.

IV. COMPLEXITY ANALYSIS

Suppose the average data size of each IIoT device is denoted as $D_{\text{avg}} = \frac{1}{N} \sum_{n \in \mathcal{N}} D_n$. We assume the time complexity of communication and computation is linearly proportional to model sizes [26]. Then, the complexity analysis of the proposed XSFL in terms of communication and computation is performed.

In traditional FL, the communication cost primarily originates from uploading the local FL models during global aggregation. As a result, the communication time complexity is $O(T|\mathbf{w}_{g,t}|)$, where $|\cdot|$ denotes the operator for calculating the size of parameters. In the proposed XSFL, the communication time complexity is $O(T|\hat{\mathbf{w}}_{g,t}^n|)$. During the communication phase, since $|\hat{\mathbf{w}}_{g,t}^n|$ is equal to $|\mathbf{w}_{n,t}|$, our methods do not directly affect the communication complexity.

The time complexity of computation which depends on the local training in the traditional FL is $O(TGD_{\text{avg}}|\mathbf{w}_{g,t}|)$, where T is the communication rounds, G is the local training iterations of each IIoT device. In XSFL, partial parameters are frozen, namely, these parameters are not trained, hence the time complexity of computation is $O(TGD_{\text{avg}}|\hat{\mathbf{w}}_{g,t}^{n,\text{tr}}|)$, where $\hat{\mathbf{w}}_{g,t}^{n,\text{tr}}$ means the trained weights of $\hat{\mathbf{w}}_{g,t}^{n}$. Due to $|\hat{\mathbf{w}}_{g,t}^{n,\text{tr}}| < |\hat{\mathbf{w}}_{g,t}^{n}|$ and the weights freezing is carried on the ES, hence $O(TGD_{\text{avg}}|\hat{\mathbf{w}}_{g,t}^{n,\text{tr}}|) < O(TGD_{\text{avg}}|\mathbf{w}_{g,t}|)$.

V. SIMULATION AND DISCUSSION

A. Simulation Settings

First, we use the Comprehensive Disaster Dataset (CDD) [27], which includes fire images, to evaluate the performance of XSFL and its competitors. The dataset is split into 80% for training and 20% for testing purposes.

Second, we consider the following contenders:

- FedAvg [11]: A benchmark FL algorithm.
- STC [28]: A compression-based FL framework tailored to FL environment requirements.
- FTTQ [29]: A ternary compression-based communication-efficient FL method.
- FedPAQ [30]: A communication-efficient FL strategy incorporating periodic averaging and quantization.
- XSFL: Our proposed FL approach for IESN, as presented in this paper.

Third, for the system model, we assume 10 IIoT devices are participating in FL training. The CPU frequency for each IIoT device is set to $f_{n,t} = 2$ GHz. Channel power gain is set at $h_{n,j}(t) = -50$ dB, and noise power is set to $\sigma =$ -174 dBm/Hz. Data transmission power and bandwidth are configured as $P^{\rm U} = 0.01$ W, $P^{\rm D} = 1$ W, $B^{\rm U} = 1$ MHz, and $B^{\rm D} = 20$ MHz [31].

B. Evaluation of FL Model Performance

In this subsection, we evaluate the proposed XSFL in terms of the model performance compared with other contenders. The evaluation indicators include Accuracy (ACC), Precision (PRE), Specificity (SPE), F1-score (F1), and Recall (REC) [32].



Fig. 3: Model performance comparison among several FL schemes.

Based on the above settings, the final evaluation results of different FL schemes are shown in the radar plots of Fig. 3. We can observe that the proposed XSFL achieves the best performance in all indicators. We speculate the excellent performance of XSFL is attributed to the ACT strategy. ACT can freeze partial weights of the global FL model for adapting to the local data volumes of the IIoT devices, which assists the local FL model in making full use of the local data and thus improves the generalization of the global FL model.

C. Evaluation of FL Training Delay

This simulation aims to demonstrate the proposed XSFL can effectively reduce the FL training delay.



Fig. 4: The average training delay of each IIoT device using several FL schemes.



Fig. 5: The values of the optimization objective for different FL schemes.

Fig. 4 shows the average FL training delay of each IIoT device in each communication round by several FL schemes. We can see that the boxplot of the presented XSFL scheme is at the bottom, which means the average training time of each IIoT device in XSFL is the shortest. Fig. 5 describes the global loss, weighted delay, and optimization objective under different FL schemes, where the weighted delay means the total training delay multiplied by a coefficient. The coefficient is set to 1e-3 to balance the delay and global loss; The optimization objective is the sum of the global loss and weighted delay. The results indicate that XSFL can effectively strike a balance between model performance and training delay in comparison to the other contenders.

D. Evaluation of Semantic Explainability

To evaluate the explainability of the XSFL, we display some ESC heatmaps corresponding to fire images from CDD.



Fig. 6: The explainable results of XSFL.

Fig. 6 displays the heatmaps corresponding to the fire images, where the key features are accurately marked, such as the fire and smoke in the fire images. Through this visual explanation, the relationship between the semantic features and input monitoring images can be captured, which helps humans interpret the prediction results and improve the trustworthiness of SC.

VI. CONCLUSIONS

In this paper, IESN is presented to reduce the communication overheads of IIoT devices for fire surveillance. Furthermore, we propose the XSFL that enables IIoT devices and ES to collaboratively train the SC model while ensuring data privacy and security. In XSFL, we first present the ACT strategy in ES, in which IIoT devices are clustered according to their local resources. Then, for each cluster, ES publishes a specific FL model that can better adapt to the device. Second, the ESC mechanism is designed to provide explainability of the fire semantics, which helps humans understand the semantic results and improve the trustworthiness of SC. Finally, the simulation results demonstrate the superiority of the proposed XSFL method.

REFERENCES

- B. Prabha, "An iot based efficient fire supervision monitoring and alerting system," in 2019 Third International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC). IEEE, 2019, pp. 414–419.
- [2] T. M. Thekkil and N. Prabakaran, "Real-time wsn based early flood detection and control monitoring system," in 2017 International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT). IEEE, 2017, pp. 1709–1713.
- [3] F. Jiang, K. Wang, L. Dong, C. Pan, W. Xu, and K. Yang, "AI driven heterogeneous mec system with uav assistance for dynamic environment: Challenges and solutions," *IEEE Network*, vol. 35, no. 1, pp. 400–408, 2020.
- [4] K. Muhammad, S. Khan, M. Elhoseny, S. H. Ahmed, and S. W. Baik, "Efficient fire detection for uncertain surveillance environment," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 5, pp. 3113–3122, 2019.
- [5] A. Jain and A. Srivastava, "Privacy-preserving efficient fire detection system for indoor surveillance," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 5, pp. 3043–3054, 2022.

- [6] G. Manikandan, D. Karunkuzhali, D. Geetha, and V. Kavitha, "Design of an iot approach for security surveillance system for industrial process monitoring using raspberry-pi," in *AIP Conference Proceedings*, vol. 2519, no. 1. AIP Publishing LLC, 2022, p. 030024.
- [7] S. Banka, B. Kanchanapalli, B. Annam, A. Akula, S. Thirukkovalluru, and S. Jarpula, "A novel method for fire detection and alarming notification using iot technology," in 2023 Second International Conference on Electronics and Renewable Systems (ICEARS), 2023, pp. 560–565.
- [8] F. Jiang, Y. Peng, L. Dong, K. Wang, K. Yang, C. Pan, and X. You, "Large ai model-based semantic communications," *IEEE Wireless Communications*, vol. 31, no. 3, pp. 68–75, 2024.
- [9] F. Yang, Y. Qiao, M. Z. Abedin, and C. Huang, "Privacy-preserved credit data sharing integrating blockchain and federated learning for industrial 4.0," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 12, pp. 8755–8764, 2022.
- [10] D. Chen, Z. Liu, L. Wang, M. Dou, J. Chen, and H. Li, "Natural disaster monitoring with wireless sensor networks: A case study of dataintensive applications upon low-cost scalable systems," *Mobile Networks* and Applications, vol. 18, no. 5, pp. 651–663, 2013.
- [11] Q. Yang, Y. Liu, Y. Cheng, Y. Kang, T. Chen, and H. Yu, "Federated learning," *Synthesis Lectures on Artificial Intelligence and Machine Learning*, vol. 13, no. 3, pp. 1–207, 2019.
- [12] E. Begoli, T. Bhattacharya, and D. Kusnezov, "The need for uncertainty quantification in machine-assisted medical decision making," *Nature Machine Intelligence*, vol. 1, no. 1, pp. 20–23, 2019.
- [13] Y. Wang, Z. Gao, D. Zheng, S. Chen, D. Gunduz, and H. V. Poor, "Transformer-empowered 6G intelligent networks: From massive mimo processing to semantic communication," *IEEE Wireless Communications*, 2022.
- [14] S. Park, O. Simeone, and J. Kang, "End-to-end fast training of communication links without a channel model via online meta-learning," in 2020 IEEE 21st International Workshop on Signal Processing Advances in Wireless Communications (SPAWC). IEEE, 2020, pp. 1–5.
- [15] L. Dong, F. Jiang, M. Wang, Y. Peng, and X. Li, "Deep progressive reinforcement learning-based flexible resource scheduling framework for irs and uav-assisted mec system," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–13, 2024.
- [16] D. Gündüz, Z. Qin, I. E. Aguerri, H. S. Dhillon, Z. Yang, A. Yener, K. K. Wong, and C.-B. Chae, "Beyond transmitting bits: Context, semantics, and task-oriented communications," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 1, pp. 5–41, 2022.
- [17] K. Wei, J. Li, C. Ma, M. Ding, C. Chen, S. Jin, Z. Han, and H. V. Poor, "Low-latency federated learning over wireless channels with differential privacy," *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 1, pp. 290–307, 2021.
- [18] X. Zhou, X. Ye, K. I.-K. Wang, W. Liang, N. K. C. Nair, S. Shimizu, Z. Yan, and Q. Jin, "Hierarchical federated learning with social context clustering-based participant selection for internet of medical things applications," *IEEE Transactions on Computational Social Systems*, vol. 10, no. 4, pp. 1742–1751, 2023.
- [19] P. Arora, S. Varshney et al., "Analysis of k-means and k-medoids algorithm for big data," Procedia Computer Science, vol. 78, pp. 507– 512, 2016.
- [20] N. Shoham, T. Avidor, A. Keren, N. Israel, D. Benditkis, L. Mor-Yosef, and I. Zeitak, "Overcoming forgetting in federated learning on non-iid data," arXiv preprint arXiv:1910.07796, 2019.
- [21] L. F. Isikdogan, B. V. Nayak, C.-T. Wu, J. P. Moreira, S. Rao, and G. Michael, "Semifreddonets: Partially frozen neural networks for efficient computer vision systems," in *European Conference on Computer Vision.* Springer, 2020, pp. 193–208.
- [22] X. Zhou, X. Zheng, X. Cui, J. Shi, W. Liang, Z. Yan, L. T. Yang, S. Shimizu, and K. I.-K. Wang, "Digital twin enhanced federated reinforcement learning with lightweight knowledge distillation in mobile networks," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 10, pp. 3191–3211, 2023.
- [23] H. Liu, Y. Wang, W. Fan, X. Liu, Y. Li, S. Jain, Y. Liu, A. K. Jain, and J. Tang, "Trustworthy ai: A computational perspective," arXiv preprint arXiv:2107.06641, 2021.
- [24] R. R. Selvaraju, A. Das, R. Vedantam, M. Cogswell, D. Parikh, and D. Batra, "Grad-cam: Why did you say that?" arXiv preprint arXiv:1611.07450, 2016.
- [25] J. Xu, Z. Li, B. Du, M. Zhang, and J. Liu, "Reluplex made more practical: Leaky relu," in 2020 IEEE Symposium on Computers and Communications (ISCC), 2020, pp. 1–7.
- [26] C. Wu, F. Wu, L. Lyu, Y. Huang, and X. Xie, "Communication-efficient federated learning via knowledge distillation," *Nature communications*, vol. 13, no. 1, pp. 1–8, 2022.

- [27] A. Omar, S. Ashraf, A. Rahman, M. A. Amin, A. A. Ali *et al.*, "A comparative study on disaster detection from social media images using deep learning," in *Proceedings of the Global AI Congress 2019*. Springer, 2020, pp. 485–499.
- [28] F. Sattler, S. Wiedemann, K.-R. Müller, and W. Samek, "Robust and communication-efficient federated learning from non-iid data," *IEEE transactions on neural networks and learning systems*, vol. 31, no. 9, pp. 3400–3413, 2019.
- [29] J. Xu, W. Du, Y. Jin, W. He, and R. Cheng, "Ternary compression for communication-efficient federated learning," *IEEE Transactions on Neural Networks and Learning Systems*, 2020.
- [30] A. Reisizadeh, A. Mokhtari, H. Hassani, A. Jadbabaie, and R. Pedarsani, "Fedpaq: A communication-efficient federated learning method with periodic averaging and quantization," in *International Conference on Artificial Intelligence and Statistics*. PMLR, 2020, pp. 2021–2031.
- [31] M. Chen, Z. Yang, W. Saad, C. Yin, H. V. Poor, and S. Cui, "A joint learning and communications framework for federated learning over wireless networks," *IEEE Transactions on Wireless Communications*, vol. 20, no. 1, pp. 269–283, 2020.
- [32] Z.-S. Chen, J.-Y. Chen, Y.-H. Chen, Y. Yang, L. Jin, E. Herrera-Viedma, and W. Pedrycz, "Large-group failure mode and effects analysis for risk management of angle grinders in the construction industry," *Information Fusion*, vol. 97, p. 101803, 2023.

BIOGRAPHIES



Li Dong received the B.S. and M.S. degrees in School of Physics and Electronics from Hunan Normal University, China, in 2004 and 2007, respectively. She received her Ph.D. degree in School of Geosciences and Info-physics from the Central South University, China, in 2018. She is currently an Associate Professor with the Hunan University of Technology and Business, Changsha. Her research interests include federated learning, Internet of Things, and mobile edge computing.



Yubo Peng received his B.S. and M.S. degrees from Hunan Normal University, Changsha, China, in 2019 and 2024. He is pursuing a doctor's degree from the School of Intelligent Software and Engineering at Nanjing University. His main research interests include semantic communication and large models.



Feibo Jiang received his B.S. and M.S. degrees in School of Physics and Electronics from Hunan Normal University, China, in 2004 and 2007, respectively. He received his Ph.D. degree in School of Geosciences and Info-physics from the Central South University, China, in 2014. He is currently an associate professor at the Hunan Provincial Key Laboratory of Intelligent Computing and Language Information Processing, Hunan Normal University, China. His research interests include semantic communication, federated learning, Internet of Things,

and mobile edge computing.



Kezhi Wang received the Ph.D. degree in Engineering from the University of Warwick, U.K. He was with the University of Essex and Northumbria University, U.K. Currently, he is a Senior Lecturer with the Department of Computer Science, Brunel University London, U.K. His research interests include wireless communications, mobile edge computing, and machine learning.



Kun Yang received his PhD from the Department of Electronic & Electrical Engineering of University College London (UCL), UK. He is currently a Chair Professor in the School of Intelligent Software and Engineering, Nanjing University, China. He is also an affiliated professor of University of Essex, UK. His main research interests include wireless networks and communications, communicationcomputing cooperation, and new AI (artificial intelligence) for wireless. He has published 400+ papers and filed 30 patents. He serves on the editorial

boards of a number of IEEE journals (e.g., IEEE WCM, TVT, TNB). He is a Deputy Editor-in-Chief of IET Smart Cities Journal. He has been a Judge of GSMA GLOMO Award at World Mobile Congress – Barcelona since 2019. He was a Distinguished Lecturer of IEEE ComSoc (2020-2021). He is a Member of Academia Europaea (MAE), a Fellow of IEEE, a Fellow of IET and a Distinguished Member of ACM.