The following publication Li, X., Chi, H. L., Lu, W., Xue, F., Zeng, J., & Li, C. Z. (2021). Federated transfer learning enabled smart work packaging for preserving personal image information of construction worker. Automation in Construction, 128, 103738 is available at https://doi.org/10.1016/j.autcon.2021.103738.

Federated Transfer Learning enabled Smart Work Packaging for Preserving the Personal Image Information of Construction Worker

Abstract

The rapidly expanding number of IoT-based camera devices makes smart work packaging (SWP) easier to access massive construction workers' personal image information for occupational health and safety (OHS) status monitoring. SWP can then transmit these personal data to the cloud for training the machine learning models and offer safety alerts or health insights. However, there are two urgently important challenges. Firstly, the machine learning model needs to aggregate the SWPs' image data from each construction worker, which may pose a risk to private data leakage without strict privacy and security agreement. In addition, the machine learning models trained on all SWPs' image data may compromise the personalization of image-based OHS status monitoring for each construction worker. To address the above issues, this study proposes a FedSWP framework, the federated transfer learning-enabled SWP for protecting the personal image information of construction workers in OHS management. FedSWP executes the gradient parameters aggregation through federated learning for the image data in each SWP and builds relatively personalized models by transfer learning. Crane operators' facial fatigue monitoring experiments are conducted and have evaluated that FedSWP can achieve accurate and personalized safety alerts and healthcare. This study paves the way for the generalization and extension of FedSWP in many construction OHS applications.

Keywords:

Federated Learning, Smart Work Packaging, Occupational Health and Safety, Transfer Learning, Privacy and Security, Image Data, Facial Fatigue

1. Introduction

Globally, the records of occupational health and safety in the construction industry are among the poorest compared to other industries (Li et al., 2018). Construction tasks are known to be executed in hazardous environments and unhealthy working conditions (Hasanzadeh et al., 2017). Despite following strict OHS regulations or transferring most site jobs to the prefabrication plant, there is no significant decline in the number of work-related injuries, illnesses, and diseases (Niu et al., 2019). Unsafe behavior, frequently located in the high-risk area, disordered biosignals (e.g., heartbeat) from construction workers are indicated as the common causes of OHS issues in the construction site (Guo et al. 2016; Fang et al. 2020). Thus, one of the most critical tasks is to track construction workers' OHS data, such as unsafe behavior motions, locations, and other biosignals in the construction workplace. Also, there is a lack of a general model for each construction worker to link their own OHS status with task executions for satisfying overall project performance. Smart work packaging (SWP) is a potential approach to facilitate construction workers preparing well by monitoring and predicting their OHS status before executing tasks (Li et al. 2019a; Li et al. 2019b; Li et al., 2020b). Recently, the expanding deployment of IoT-based camera devices mounted in the construction worker's personal protection equipment (PPE) or the workface of construction sites improve the capacity of SWP in sensing and tracking image information of each construction worker for ergonomic posture recognition, physical fatigue assessment, falls and unsafe behavior detection (Yan et al. 2017; Yu et al. 2019; Fang et al. 2018). Meanwhile, the rapid development of machine learning techniques enhances the SWP to help construction workers to understand their OHS status by processing, networking, and reasoning their image information into each construction workflow. SWP with deep learning and optimization techniques has also been proved to offer early warnings or predictions to OHS issues such as facial fatigue (Li et al., 2019a) and unsafe behavior (Li et al., 2020b).

Image data-processing models in construction task executions, such as machine learning models embed in SWP, are usually trained on sufficient image data to monitor and predict construction workers' OHS status. The traditional image data-processing mechanism in SWP involves simple image data aggregation models, in which cameras capture video data to the cloud, and the cloud is responsible for cleaning, merging and transferring them into the image data. Finally, the cloud will use the aggregated image data and establish standard models for OHS applications. However, there are two critical challenges in the image data-processing mechanism of SWP in construction OHS. **Firstly,** image data for unsafe behavior or motions, facial fatigue, disease and medical reports are sensitive and private. Thus, image datasets for construction OHS are hard to be collected, and they keep in distributed construction workers or isolated construction sites. The insufficient image data sources can result in unsatisfied machine-learning models' performance, which serves as the critical constraint of SWP for construction OHS management. Moreover, increasing numbers of countries enforce laws or regulations to protect data privacy and security, such as the General Data Protection Regulation (GDPR) published by the European Union (Voigt et al., 2017). **Secondly,** machine learning models in SWP lack personalization for each construction worker. Current methods rely on a standard cloud model for monitoring the OHS statuses of all construction workers. After aggregating adequate image data to get a satisfied machine learning model, then the model is distributed to all SWPs. However, different construction workers have various physical characteristics and daily work patterns. Thus, the standard model in SWP can not perform personalized OHS management (Chen et al., 2020).

Thus there is a dilemma that construction image data is in the distributed SWP and isolated construction sites. Still, it is prohibited from collecting, aggregating, and using the image data for machine learning processing. How to legally and ethically solve image data usage and isolation in SWP is a significant challenge for both construction professionals and academicians. Federated transfer learning could be a potential solution for these challenges (Yang et al., 2019).

This study aims to propose FedSWP, the federated transfer learning-enabled SWP framework for preserving personal image information of construction workers to achieve better OHS management, improving data isolation and personalization issues of SWP. To satisfy this aim, three concrete objectives are explained below: (1) To establish a federated learning model and define related SWP workflow for image data privacy persevering; (2) To utilize transfer learning for achieving personalized model learning for each SWP; (3) To validate the FedSWP in a scenario of crane operator facial fatigue monitoring. The contributions to the body of knowledge in this study can be summarized as threefolds. First, to the authors' knowledge, this study is one of the first investigations on federated transfer learning in the construction OHS to improve the image data privacy-preserving for each construction worker at the work package level. To this end, a framework of FedSWP with five tasks is proposed. Second, the hybrid deep neural network in FedSWP is developed and customized to achieve better accuracy and personalization in the complex tasks of facial fatigue monitoring and prediction, which has very sensitive facial expressions and very dynamic spatial-temporal features. Third, FedSWP ensures the task of facial fatigue prediction by gathering encrypted model parameters instead of directly capturing the private facial image, keeping the training dataset locally, and ensuring privacy-preserving of the raw data. The rest of the paper is organized as follows. Subsequent to this introductory section is Section 2 to elaborate backgrounds on construction OHS, SWP, and federated transfer learning.

Section 3 is to delineates a FedSWP framework, followed by the experiment elaborated in Section 4. Section 5 is a discussion to present the novelties and limitations of this study. Conclusions are drawn in Section 6.

2. Background

2.1 Construction Occupational Health and Safety

Construction on-site activities cover intensive workloads and are physically demanding in hazardous and unhealthy environments, which result in high rates of injuries, fatalities, musculoskeletal, and cardiovascular diseases (Lee et al., 2017). Previous studies have proved that most of these OHS issues do not occur at random when reaching high monitoring and predictive skills (Hallowell et al. 2013; Tixier et al. 2016; Park and Kim, 2013). Thus, numerous tracking and sensing technologies (See Fig.1) have been empirically and quantitatively used to monitor OHS status for construction workers rather than being approached by analyzing subjective data and expert opinion. As shown in Fig.1, numerous IoT sensors can be embedded in the personal protective equipment (PPE) for invasive-free OHS status monitoring. For example, smart helmets with GPS can help track workers whether frequently appeared in the hazardous workplace (e.g., crane operation area) (Edirisinghe, 2019). Smart glasses with cameras can facilitate monitoring of workers' fatigue or unsafe behaviors through image process techniques (Chang et al., 2018). Smart vests with sensors of motion, temperature, humidity, and other biomedical pads can sense heat stress, muscle strain, or other cardiovascular signals (Ahn et al., 2019). However, previous studies were limited to exploring a general model for managing isolated OHS data and making the OHS insights (e.g., predictions, warnings) ready for each distributed construction worker before task executions at the work package level. The smark work packaging (SWP) is such an approach to helping model, optimize, and monitor each worker's OHS status and facilitate their task executions.

Figure 1. OHS sensory data for construction worker

2.2 Smart Work Packaging

SWP is defined as an approach to decompose the construction workflows and integrate smartness capabilities, such as visualizing, processing, networking, and reasoning into the workflows so that they can be executed autonomously, adapt to changes in their physical context, and interact with the surroundings to enable the more resilient process (Li et al., 2019b). Numerous techniques have been used to model, optimize, and monitor the OHS data in SWP for each construction worker. For example, the hybrid system dynamics-discrete event simulation model in SWP can help assess the impacts of each workers' OHS constraints on overall project schedule performance (Li et al., 2019c). In addition, the probabilistic roadmap model serves in SWP to facilitate task optimization by capturing dynamic OHS data (e.g., locations of walking workers) (Li et al., 2020). Furthermore, the hybrid deep neural networks (convolutional neural networks (CNN) and bidirectional long-short term memory (Bi-LSTM)) developed in SWP have been used efficiently to monitor and predict the crane operators’ fatigue (Li et al., 2019a; Li, 2019). Instead of introducing an entirely new workflow in monitoring the worker’s fatigue, SWP augments existing workflows with smart characteristics, including adaptivity, sociability, and autonomy (Li, 2019). The crane operator fatigue monitoring and alerting needed in the crane operations is the trial to activate the potential of proactive tracking, updating, and predicting in SWP’s autonomy. The SWP-enabled fatigue monitoring service has been investigated in the author’s previous study, and the architecture of this service is shown in Fig.2 (Li et al., 2019c).

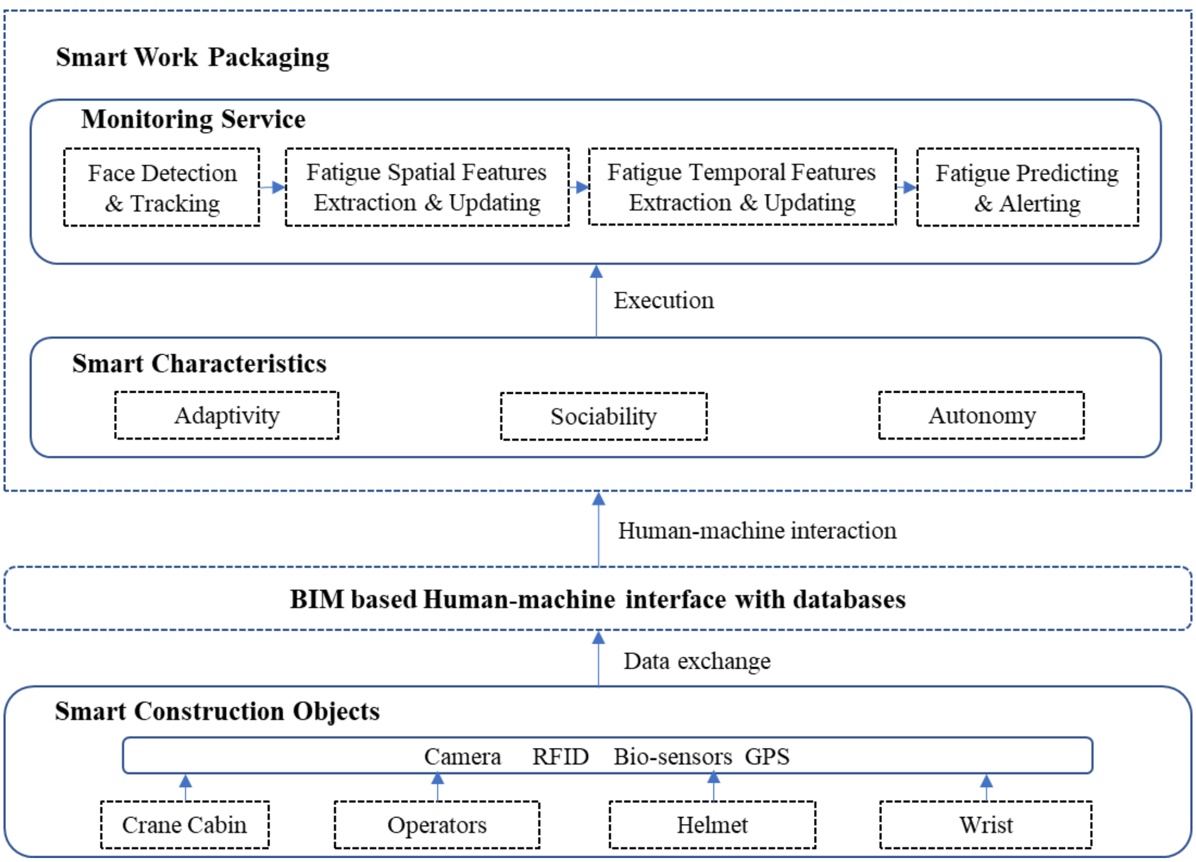

Figure 2. The architecture of SWP-enabled fatigue monitoring service

This service is supported by IoT-based construction resources (smart construction objects (SCOs)) and the smart BIM platform. The SCOs are built by equipping the objects such as crane cabin, operator, helmet and wrist on the operator with the various sensing and tracking technologies (e.g., RFID, bio-sensors for monitoring fatigue, WiFi, camera) for achieving smartness in data generation and collection. This process can both enrich and exchange information with the smart BIM platform. Then, after the interactions between the crane operators and the human-machine interface, SWP can be activated to execute the tasks of fatigue monitoring. Models and techniques in SWP often need to share or aggregate data from each construction worker. However, the use of workers' personal data may generate risks to privacy issues. The privacy concerns raised by construction workers can impede the dissemination of SWP and related wearable technologies in construction sites. For example, SWP for OHS management has a good intention to prevent workers' access to a hazardous construction environment by using location-based technologies. However, construction workers may be reluctant to reveal their location data, as they think the data can be used for surveilling their idling time (Choi et al., 2017) (Seo et al., 2015). IoT Camera-based image process techniques in SWP are also most widely used for construction OHS monitoring (Mostafa and Hegazy, 2021). For example, Tang et al. (2020) improved the breadth and depth of vision-based safety compliance checking by explicitly classifying worker-tool interactions. Wu et al. (2019) proposed a one-stage system based on a convolutional neural network (CNN) to automatically monitor whether construction workers are wearing hardhats or not. Luo et al. (2019) integrated the latest computer vision methods to detect and visualize the dynamic workspaces of construction workers on foot. Son et al. (2019) used the deep residual networks to detect construction workers under varying poses and against a dynamic environment. Ding et al. (2018) also developed a hybrid deep learning model to detect unsafe behavior with complex Spatio-temporal features. These studies have frequently involved private human-related images. Thus, technical solutions for privacy-preserving SWP to relieve workers' privacy concerns are urgently needed, particularly image information. Federated learning holds the key.

2.3 Federated Transfer Learning

Federated learning is firstly proposed by Google to keep data trained locally on the distributed users' mobile devices and updated parameters to a global machine learning model, which aims to protect user data privacy (Konečný et al.,2016). Since then, numerous studies push forward the evolution of federated learning, such as optimization improvement (e.g., reduce communication cost, heterogeneity) (Sattler et al., 2019), security analysis (e.g., blockchain-enabled privacy-preserving technologies) (Lu et al., 2019), and expanding applications (e.g., mobile devices, industrial engineering, healthcare) (Li et al., 2020a). SWP assigned to each construction worker can be considered as the distributed multi-agent network. Thus, federated learning can resolve data islanding and privacy issues by training machine learning models in the SWP network. Yang et al. (2019) classified federated learning into three categories: (1) horizontal federated learning, where distributed parties share similar features but differ in samples; (2) vertical federated learning, where distributed parties share similar samples but vary in features; (3) federated transfer learning, where distributed parties not only differ in samples but also in features. As OHS data of construction workers in each SWP is from different samples and may share a few features, FedSWP can be the federated transfer learning (FTL). It would be the first customized FTL model for construction OHS. Like the human's learning processes on cross-domain knowledge, transfer learning can employ existing knowledge from a familiar domain to improve the learning performance or minimize the training processes in a new domain (Zhuang et al., 2020). Based on the distribution variation between different domains, transferring learning is classified into homogeneous and heterogeneous (Weiss et al., 2016). The former indicates the domains have the same features and accustom the domains by adjusting the sample selection bias (Bickel et al., 2009). The latter presents the domains with different features, requiring feature adaption (Long et al., 2013). As the availability of labeled OHS data is unrealistic, FedSWP mainly used deep transfer learning models in the federated learning paradigm.

3. The Proposed FedSWP Framework

3.1 Problem Definition

In this study, the construction workers are the entities in facial fatigue status monitoring and prediction. SWPs are the computing nodes that each SWP corresponds to each worker and includes the sensory camera. In the context of facial fatigue status monitoring and prediction, construction workers can be considered as SWPs in the proposed FedSWP framework. All SWPs have image datasets D. FedSWP aims to help monitor and predict the facial fatigue status with historical image data from isolated SWPs without sharing any data and data privacy leakage. Given $W = \{W_1, W_2, \cdots W_N\}$ denotes the SWPs (N workers) set, and the datasets in SWPs can be denoted by $D = \{D_1, D_2, \cdots D_N\}$. Let t and O_t denote the t-th timestamp for temporal data and fatigue status at the t-th timestamp. Let $f(t, D)$ be the function of facial fatigue status prediction, the problem definitions of personalization and privacy can be summarized as follows:

Privacy: this study defines privacy as avoiding access to private facial image data, which may be closely related to the construction worker's personal information. For example, a camera mounted in the crane cabin monitors operator fatigue status and allows the project manager to access crane operators' facial or other behaviors, violating the privacy definition. In this study, each SWP trains its local model on local datasets rather than sharing any data and only submitting the local model's updated parameters to the cloud.

Personalization: this study defines personalization as higher accuracy in predicting facial fatigue status for each SWP. Previous methods for facial fatigue status monitoring and prediction aggregate all the datasets $D = D_1 \cup D_2 \cup \cdots \cup D_N$ to train a machine learning model and calculate $O_{t+s} = f(t+s, D)$ with the accuracy of A_{ALL}, where s is the prediction window after t. As all datasets' distributions may vary and OHS datasets in SWPs require privacy-preserving. The federated transfer learning model for each SWP trained as $O_{t+s} = f_i(t+s, D_i)$ and the accuracy can be denoted by A_{FED}. The hypothesis of FedSWP is then to test whether the accuracy of federated transfer learning is better than traditional methods, which can be denoted by

$$A_{FED} - A_{ALL} > 0 \qquad (1)$$

3.2 Framework Overview

This section proposes the FedSWP framework, which enables SWP for privacy-preserving and personalized fatigue monitoring by employing federated transfer learning. Fig.2 presents the FedSWP framework and five tasks included for each SWP. Firstly, a machine learning model is selected and trained on the public datasets to get a global model. Secondly, this initial global model is shared with each SWP. Thirdly, the local model is trained on each worker's image database. In addition, SWP updates the local models' parameters to create a new global model. It should be noted that this task sends back the local model's encrypted parameters rather than sharing any construction workers' facial image data. Finally, each SWP can get the personalized model by performing transfer learning from the global model to the local model.

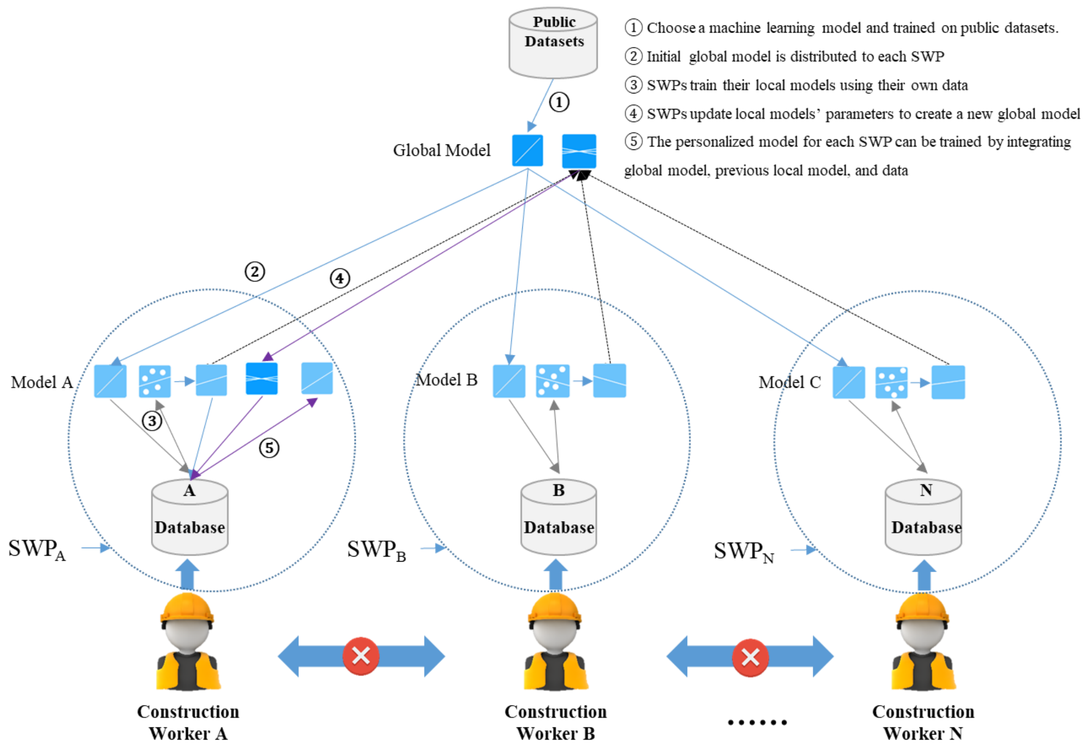

Figure 2 FedSWP framework for federated transfer learning

3.3 Federated Transfer Learning

FedSWP takes the federated transfer learning paradigm (Chen et al. 2020; Liu et al. 2020) to achieve encrypted parameter sharing and personalized model training. According to the above framework, federated transfer learning mainly comprises two critical parts: federated learning and transfer learning. After receiving the global machine learning model, it is distributed to SWP for federated learning by training the local model and obtaining the updated global model. Then the updated global model is integrated with each local model for transfer learning.

3.3.1 Federated Learning

In FedSWP, deep neural networks are adopted to train the global and local models. It can assume for fatigue monitoring and prediction task in each SWP has one database, and the local training dataset can be defined as $D = \sum_{n=1}^{N} D_n$. In the typical deep neural networks, given the sample set

$\{x_i, y_i\}_{i=1}^{D_n}$, where the input vector is $x_i \in \mathbb{R}^d$ with d dimension of features, and the output value is $y_i \in \mathbb{R}$. The model parameter vector $\omega \in \mathbb{R}^d$ (e.g., weight and bias) can be obtained by fitting the output y_i with loss function $\ell\,(\cdot,\cdot)$, e.g., mean square error (MSE) for the regression model. Let f_G denote the global model to be learned, and the learning objective can be represented as follows:

$$\arg\min_{\omega \in \mathbb{R}^d} \mathcal{L}(\omega) = \sum_{i=1}^{D_n} \ell(f_G(x_i), y_i) \qquad (2)$$

Once receiving the global model, it can be broadcast to all SWPs via additively homomorphic encryption for local training. When sending back the well-trained parameters, homomorphic encryption can also prevent parameter leakage in the parameter sharing processes (Aono et al., 2017). Let $E(\cdot)$ represent the homomorphic encryption function, for any two parameters ω_1 and ω_2 in two different SWPs, the additively homomorphic encryption can be achieved as follows:

$$E(\omega_1) + E(\omega_2) = E(\omega_1 + \omega_2) \qquad (3)$$

Similarly, let f_W denote the local model to be learned in SWP, and the learning objective can be represented as follows:

$$\arg\min_{\omega^W \in \mathbb{R}^d} \mathcal{L}(\omega^W) = \sum_{i=1}^{{D_n}^W} \ell(f_w(x_i), y_i) \qquad (4)$$

When all local models f_W are trained, they should be uploaded for aggregation. However, the limited communication bandwidth can lead to high-latency and low-throughput for aggregating local model's updates from SWPs. To mitigate the communication cost, each SWP can implement gradient descent optimization based on its local dataset. The cloud then conducts weighted average aggregation on their updates from the SWPs. See Algorithm 1, the federated averaging algorithm contains three steps: (1) randomly select subsets of SWPs and broadcast the global model to the selected SWPs; (2) Each SWP W trains data locally and updates ω_t for E epochs of stochastic gradient descent (SGD) (e.g., Adam) with mini-batch size to get $\omega_{t+1,w}$; (3) The cloud aggregates each SWP's ω_{t+1} through additively homomorphic encryption.

Algorithm 1 Federated Averaging Algorithm

Algorithm 1: Federated Averaging Algorithm

Input: SWP $W = \{W_1, W_2, \cdots W_N\}$, C is the selection proportion of SWPs in each Round, B is the local minibatch size, E is the number of local epochs, and η is the learning rate

Output: Parameter ω

Cloud Execution:

Initialize ω_0 (Pre-trained by a public dataset)

for each round t = 1, 2, … **do**

S_t ←Random subset of $\max(C \cdot W, 1)$ SWPs in the current round

for each SWP $w \in S_t$ in parallel **do**

$\omega_{t+1,w} \leftarrow$ SWPupdate (w, ω_t)

$\omega_{t+1,w} \leftarrow \sum_{w=1}^{N} \frac{n_w}{n} \omega_{t+1,w}$

SWPupdate (w, ω_t): // *Executed on SWP w*

for each epoch $e \in [1, E]$ **do**

Split local dataset in $\mathcal{B}$ ($\frac{B}{n}$ batches of size B)

for batch $b \in \mathcal{B}$ **do**

$\omega \leftarrow \omega - \eta \nabla \ell(\omega; b)$

Return ω to Cloud

3.3.2 Transfer Learning

Federated learning can help achieve the privacy-preserving for using construction workers' image data to train a global machine learning model. However, after federated learning, the global model can only impact when it can conduct facial fatigue monitoring and prediction for each worker in a personalized manner. As the dataset distribution may differ significantly between each SWP and the cloud, the performance could be poor for a specific SWP. The global model may only capture the coarse features from all SWPs' large-volume image data, but some dynamic and fine-grained features can be ignored, which may be very important for facial fatigue monitoring and prediction. In this study, FedSWP adopts transfer learning to achieve personalization in facial fatigue monitoring and prediction for each SWP. As the common features are always located in the shallow layers of deep neural networks, they can be transferred from the global model (Yosinski et al., 2014). Only the deep layer's specific features are trained and integrated with the transferred model to learn each SWP's personalized model.

Figure 3 shows the transfer learning process for a specific hybrid deep neural network designed to monitor and predict crane operators' fatigue where the input data is the private face image or video and the output is the fatigue level. This hybrid deep neural network includes a face detector (MTCNN, proposed by Zhang et al. 2016), a spatial feature extractor (MobileNet, offered by Howard et al. 2017), and the deep bidirectional long short-term memory (LSTM) for analyzing hidden sequential patterns in temporal features (Graves et al., 2013). MTCNN comprises three neural networks (P-Net, R-Net, and O-Net), and it is used to get face windows and landmarks. MobileNet includes thirteen convolutional layers (Conv 1-13), five max-pooling layers (Max Pool 1-5), one average-pooling layer (Ave Pool), and one fully connected network layer (FC), which is adopted to extract the common features on the face. Thus, MTCNN and MobileNet are frozen in the transfer learning process, which means their parameters will not be updated in backpropagation. As LSTM can learn long-term dependencies to learn high-level dynamic temporal features for SWP, its parameters will be updated with the training process. The softmax works as the activation function to normalize the output to a probability distribution over predicted output results, which can be formulated as:

$$\sigma(z)_i = \frac{e^{z_i}}{\sum_{j=1}^{C} e^{z_j}} \tag{5}$$

Where σ is softmax to get final prediction results, z is the input vector, e^{z_i} is the standard exponential function for input vector, C is the number of states, e^{z_j} is the standard exponential function for the output vector.

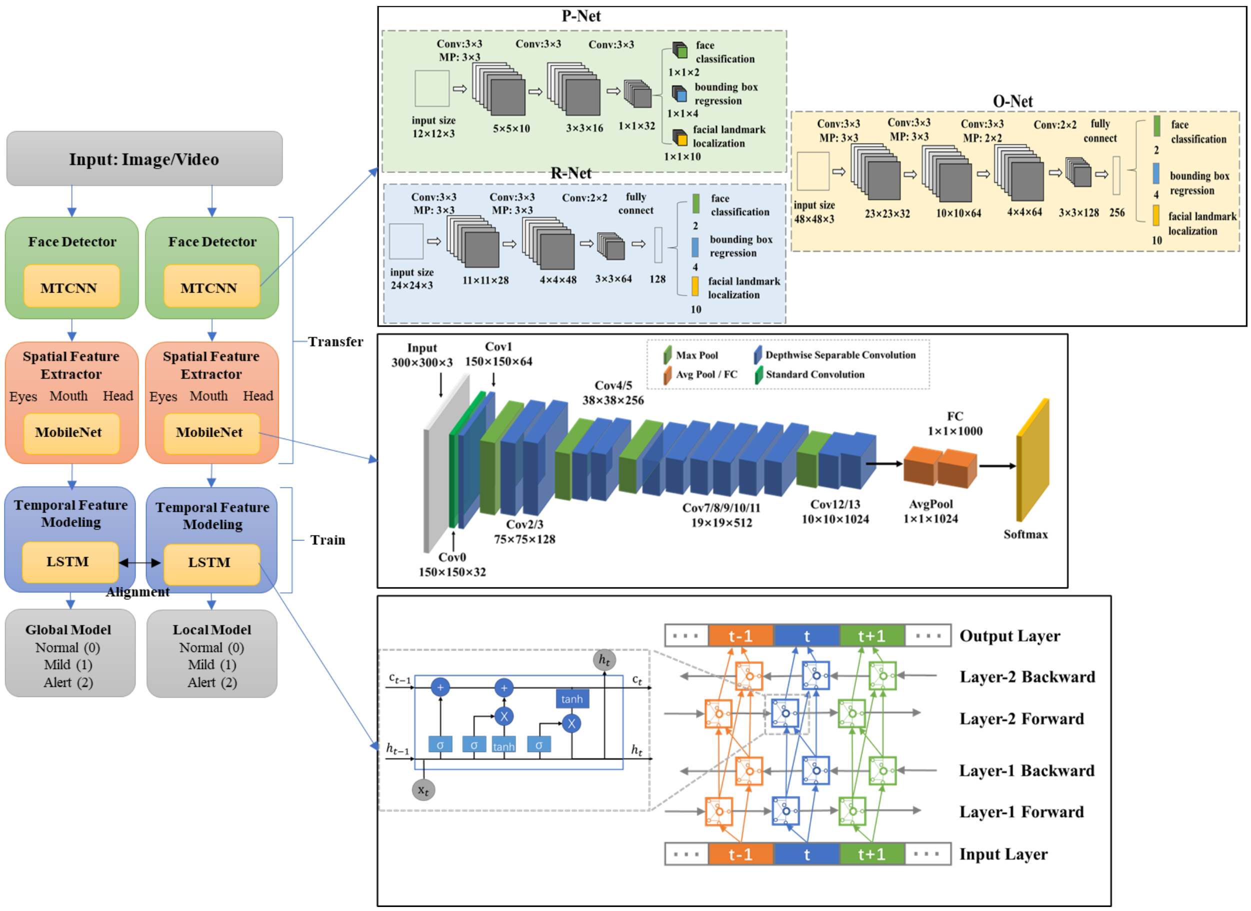

Figure 3. The transfer learning process of FedSWP

The deep neural networks cannot always generalize well across different domains from the local dataset and global model. Thus, an alignment layer can be used in FedSWP to replace the fully connected layer, which is located after LSTM and before softmax. This alignment function is correlation alignment, named CORAL (Sun et al., 2016), which can help align the second-order statistics between the source (global) and target (local) inputs. The related loss function can be represented below:

$$\ell_{CORAL} = \frac{1}{4d^2} \|C_S - C_T\|_F^2 \tag{6}$$

Where $\|\cdot\|_F^2$ demonstrates the squared matrix Frobenius norm, and d denotes embedding features dimension. C_S and C_T represent the covariance matrices of the source (global) and target (local) weights. Combined training with both the regression loss and CORAL loss could learn features that perform well on the target domain (local dataset). Thus the loss for the local model can be computed by:

$$\arg \min_{\omega^W \in \mathbb{R}^d} \mathcal{L}(\omega^W) = \sum_{i=1}^{D_n{}^W} \ell(f_w(x_i), y_i) + \lambda \ell_{CORAL} \tag{7}$$

Where λ denotes the trade-off parameter.

In summary, the global model updates itself by aligning with local models and then broadcast its updated one to all SWPs for deploying transfer learning on each SWP to obtain the personalized local model using Eq.(6).

4. Experiment

In this section, experiments are conducted to evaluate the performance of FedSWP on facial fatigue monitoring and prediction, which is one of the most private OHS monitoring tasks in construction, particularly for crane operators.

4.1 Dataset

As it is believed that monitoring and alerting of the crane operator fatigue are almost identical to the situation for the vehicle driver, evaluating the performance of this proposed FedSWP method and comparing with previous methods can be conducted on a very popular public dataset named YawDD (Abtahi et al., 2014). YawDD comprises two datasets with various facial characteristics, and they are collected by natural and varying illumination conditions with the resolution of

640×480 pixels 24-bit true color (RGB) videos. The first dataset is collected by a camera mounted in the cockpit's front mirror. Each subject includes three or four videos, and each video comprises various facial states such as normal, talking/singing, and yawning. This dataset offers 270 videos for 90 subjects (47 male and 43 female). The second dataset is collected by a camera mounted on the cockpit's dashboard. Each subject includes only one video, and this dataset comprises 29 subjects (16 male and 13 female). Both datasets include 5 scenarios, such as Bareface, Glasses, Sunglasses, Mustache, Breard, in which they were collected at 30 frames per second (fps). Each subject may have a combinational representation of fatigue-related expressions (yawning, nodding, slow blink rate of eyes) and non-fatigued related signs (talking, singing, laughing, normal stillness). Table 1 summarizes the details on YawDD. To create a FedSWP scenario, the standard-setting is adjusted on this dataset. Ten subjects' videos are randomly selected from these two datasets to train the initial global model, and three subjects' videos are considered as the local dataset in each worker (subject)'s SWP, which can not share data. This setting aims to build the global model and use all three isolated SWP video data to evaluate the performance of fatigue monitoring and prediction on the three subjects without privacy leakage.

Table 1. Descriptive information on YawDD

Dataset	**Subjects**	**Behavior**	**Illumination**	**Camera Type**	**Scenarios**	**Age**	**Camera Position**
Yaw DD	119	• Normal Stillness • Talking or Singing • Yawning • Sleepy Blinking • Nodding	Day (from early morning till sunset)	RGB	• Glasses • Sunglasses • Mustache • Beard • Bare Face	20-59	Under the front mirror of the cockpit; On the cockpit dashboard

4.2 Implementation Details

Environment Setting: The experiment, including the training and evaluation process, is conducted in two virtual machines that run CentOS 7.4 system. The specification and configuration of this computer are in the following:

- CPU: Intel (R) Xeon (R) Gold 5120 × 2
- GPU: NVIDIA GeForce GTX 1080 Ti × 4
- RAM: 16GB DDR4 Memory
- Hard Disk: 500 GB SSD and 4TB HDD
- Run-on GPU: MTCNN, MobileNet, LSTM, and GRU

The algorithms are built using Python 3.6.8 with TensorFlow version 2.3.1 and Keras version 2.4.3 as the deep learning framework. OpenCV 4.4 is also adopted as a real-time image operation cross-platform and open-source computer vision library.

Data Pre-processing: For all videos, the MTCNN is used to detect the faces and locate the frames' landmarks. The detected and aligned face with five landmark points is cropped and resized to a fixed size (64×64). The ground truth of fatigue status includes the facial expressions on the eyes, head, and mouth. It is easy to find that the ground truths in each video (each annotation file) are long-term dependencies, which indicates that the states of a frame may rely on the frames in the past or future several seconds. However, these facial expressions on a series of frames, within a few seconds, would still be considered as fatigue signs if they had just restored the expressions to alert after drowsy states (Lyu et al., 2018). Furthermore, the existing labels on this dataset cannot accurately identify the fatigue states in the temporal dimension. Those typical facial states or behaviors, such as closing eyes, yawning, and lowering head, are still considered as the evidence to judge whether a frame contributes to the awareness of fatigue. To accurately describe the transitional states between the alert and the fatigue, the datasets are relabeled into three fatigue levels: alert, low vigilant, and fatigue. (1) Normal (labeled as 0): normal means the subject is experiencing no signs of fatigue or drowsiness. (2) Mild (labeled as 1): mild indicates situations that some fatigue signs appear or present but do not last for a while. (3) Alert (labeled as 2): alert means the subject presents the biosignals of drowsiness. As shown in Table 2, the behaviors, such as stillness, looking aside, normal blinking and talking, laughing, and singing, are the least related to fatigue. Therefore, they can be relabeled to 0. To achieve early fatigue detection, behaviors like distraction and sleepy blinking are defined as the change states from normal to alert or fatigue signs. They can be relabeled to 1. As for obvious fatigue behaviors, like yawning and nodding, they can be relabeled to 2. Table 3 presents the statistical information of the two datasets in YawDD, randomly selected datasets for the global model (from YawDD), and randomly selected three individual datasets for the local model (from global model databases).

Table 2. Detailed information on YawDD

Behavior	**State**	**Fatigue Level**
Talking, laughing, singing Looking aside Normal blinking Stillness	Normal	0
Distraction Sleepy-eyes Sleepy blinking	Mild: Transitional States	1
Drowsy Yawning Nodding	Alert	2

Table 3. Statistical information on Datasets

Subject	Status	Instance	Type	Shape	Source
90	3	270	Videos	(:, 640, 480, 3)	Mirror
29	3	29	Videos	(:, 640, 480, 3)	Dashboard
10	3	38998	Array	(30, 512)	Trainable Global
1	3	2233	Array	(30, 512)	Local-P1
1	3	2374	Array	(30, 512)	Local-P2
1	3	2578	Array	(30, 512)	Local-P3

Training: On both the global and local end, the hybrid deep neural networks (MTCNN, MobileNet, LSTM) were adopted for training and prediction (See Table 4). In the federated learning process, the sequential features are computed from the eyes, mouths, head areas for each subject. From all available data obtained from video clips, 80% were randomly selected to train the model, then were evaluated using the other 20% from the remaining data. For the optimization, the Adam (Kingma and Ba, 2014), as the extension of stochastic gradient descent (SGD), is chosen with mini-batch size 64, learning rate 0.0001, clip value 5, and the trade-off parameter λ at 0.01. Additively homomorphic encryption is used in model distribution and parameter sharing during federated learning training and evaluation. In transfer learning, the MTCNN and MobilNet are frozen for transferring. Only the parameters of LSTM and its fully connected layer (Dense) are trained and updated with FedAvg. As a regularization technique, the dropout is adopted for reducing overfitting of LSTM by preventing complex co-adaptations on the training data (Gal and Ghahramani, 2016). The effectiveness of FedSWP can be evaluated from two aspects. Firstly, FedSWP with LSTM can be compared with NonFed LSTM, where performances of NonFed LSTM on each subject are recorded by using the global model (without federated learning). In addition, the performances of other machine learning models, such as gated recurrent unit (GRU) with FedSWP, can also be used for comparison. The hyperparameters, such as the window size, the number of hidden layers, are critical factors that affect the model performance. We tuned LSTM and GRU under federated learning (we named them FedLSTM and FedGRU) to explore the best performance. The optimal window size is 30, and the number of hidden layers can be in [1, 2] (we named them LSTM 1, LSTM 2, GRU 1, GRU 2).

Table 4. The detailed information on hybrid deep neural networks in FedSWP

Training process			
Type	Trainability	Model	Layers
Global model	Non-trainable	MTCNN (Face Detection)	P-Net(12,12,3)
			R-Net(24,24,3)
			O-Net(48,48,3)
	Non-trainable	MobileNet (Feature Extraction)	Input_1(:,224,224,3)
			Conv_1(:,112,112,32)
			…
			Conv_13(:,7,7,1024)
			Global_average_pooling2d_1(:,1024)
			Reshape_1(:,1,1,1024)
			Dropout(:, 1, 1, 1024)
			Conv_preds(:, 1, 1, 1000)
			Softmax(:, 1, 1, 1000)
			Dense_1(:, 1, 1, 512)
			Flatten_1(:, 512)
	Trainable	LSTM/GRU(Prediction)	LSTM/GRU_1（:, 30, 512）
			LSTM/GRU_2（:, 256）
			Dense_1(:,128)
			Dropout_1(:,128)
Local model	Trainable	LSTM/GRU(Prediction)	Dense(1)

Evaluation Metrics: The FedSWP's performance was evaluated quantitatively via the metrics of accuracy and loss (See Eq.8 and Eq.9). Accuracy refers to the percentage of the personalized dataset for each subject (each SWP) that has been predicted correctly, and all experiments are conducted five times to record the average accuracy. The loss (Mean Squared Error, MSE) is the average squared difference between the estimated and the actual value:

$$Accuracy = (TP + TN)/(TP + TN + FP + FN) \quad (8)$$

$$Loss = \sum_{i=1}^{N} (Y_i - \widehat{Y_i})^2 / N \quad (9)$$

TP, *TN*, *FP*, and *FN* represent true positive, true negative, false positive, and false negative individually, based on the comparisons between the fatigue prediction results and ground truths. Y_i denotes the fatigue level being predicted and $\widehat{Y_i}$ represents the actual label value. N is the number of video frames being used for fatigue detection.

4.3 Results Analysis

4.3.1 Prediction Accuracy

Figure 4 compares NonFed and FedSWP for three subjects' facial fatigue prediction task, which shows the accuracies of facial fatigue prediction for each subject (P1, P2, P3). Compared with the authors' previous study (Liu et al., 2020), the accuracy of NonFed (centralized LSTM with two hidden layers (LSTM 2)) (P1: 66.41%, P2: 68.48%, P3: 44.46%) is much lower than the previous results on YawDD (87.52%). The underlying reason could be that the previous study's model is trained and tested on the YawDD, while NonFed as a complex model trained on the global dataset (only 10 subjects) must be overfitted and tested with poor performance on the local dataset. The proposed FedSWP with both LSTM and GRU achieves better prediction accuracy than NonFed on three subjects. Compared to NonFed, it evidently improves the average accuracy by 26.68% (FedLSTM 2) and 26.97% (FedGRU 2).

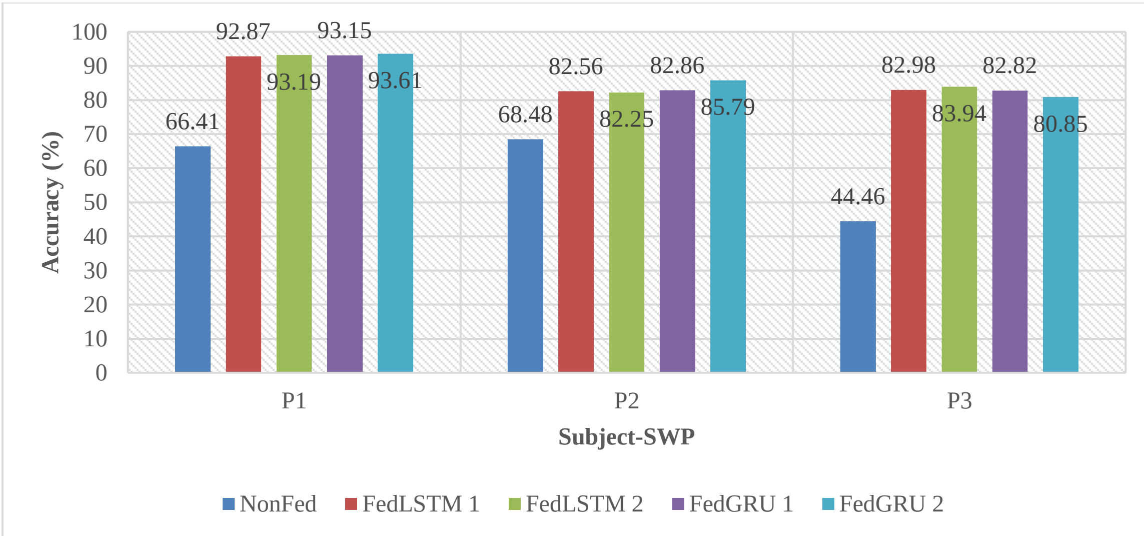

Figure 4. Comparison of fatigue prediction accuracy between NonFed, FedLSTM 1 (one hidden layer), FedLSTM 2 (two hidden layers), FedGRU 1 (one hidden layer), and FedGRU 2 (two hidden layers)

The traditional machine learning methods for general prediction applications (e.g., support vector machine (SVM)) have been proved to underperform with temporal data in fatigue prediction (Li et al., 2019a; Li et al., 2019d). However, GRU is more comparable to LSTM as they have similar recurrent cell designs for modeling temporal data (Cho et al., 2014). Thus, we compare the performance of LSTM in FedSWP (FedLSTM) with GRU in FedSWP (FedGRU). This study also evaluates the performance of FedLSTM and FedGRU through several metrics, such as Mean Absolute Error (MAE), Mean Square Error (MSE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE). The optimal architecture of FedLSTM and FedRGU comprises two hidden layers, as shown in Table 5-7. For P1 and P2, FedGRU outperforms FedLSTM both for the models with two different hidden layers. For P3, FedLSTM predicts better than FedGRU. However, the average differences between FedLSTM and FedGRU are no more than 3%. The above comparisons can also be proved in the visualized fatigue prediction process for P1 (See Fig.5). In Fig.5, FedLSTM 2 keeps more aligned with ground truth than NonFed, and FedLSTM2 is comparable to FedGRU 2. In general, the prediction accuracy results indicate the effectiveness of the personalization in the FedSWP when comparing FedLSTM/FedGRU with NonFed. It is interesting to find that the accuracy performance of FedLSTM is very close to that of FedGRU. The underlying reason could be that the basic structure of FedLSTM is similar to FedGRU. In summary, the accuracy performance of FedLSTM can be comparable to the FedGRU.

Table 5. Prediction Results of P1

Metrics	Hidden Layer	Accuracy	MAE	MSE	RMSE	MAPE
FedLSTM	1, (256)	92.87 %	0.1020	0.1634	0.4043	0.1313
	2, (512, 256)	93.19 %	0.1006	0.1658	0.4072	0.0227
FedGRU	1, (256)	93.15 %	0.1006	0.1648	0.4060	0.0449
	2, (512, 256)	93.61 %	0.0898	0.1420	0.3769	0.0686

Table 6. Prediction Results of P2

Metrics	Hidden Layer	Accuracy	MAE	MSE	RMSE	MAPE
FedLSTM	1, (256)	82.56 %	0.3194	0.6097	0.7808	0.1411
	2, (512, 256)	82.25 %	0.3295	0.6337	0.7960	0.0441
FedGRU	1, (256)	82.86 %	0.3046	0.5712	0.7558	0.2857
	2, (512, 256)	85.79 %	0.2486	0.4619	0.6796	0.1516

Table 7. Prediction Results of P3

Metrics	Hidden Layer	Accuracy	MAE	MSE	RMSE	MAPE
FedLSTM	1, (256)	82.98 %	0.2532	0.4201	0.6481	0.4927
	2, (512, 256)	83.94 %	0.2427	0.4073	0.6382	0.4637
FedGRU	1, (256)	82.82 %	0.2632	0.4462	0.6680	0.4979
	2, (512, 256)	80.85 %	0.3049	0.5321	0.7294	0.5372

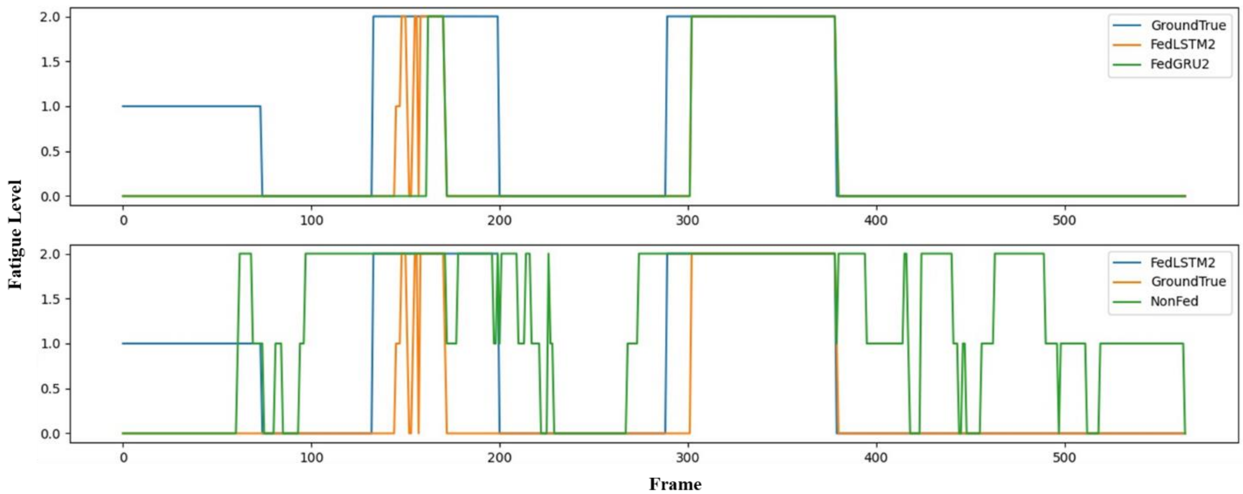

Figure 5. Fatigue prediction process for P1 by NonFed, FedLSTM 2, and FedGRU 2

Figure 6 illustrates the loss curves of the baseline NonFed LSTM model and FedSWP models (FedLSTM and FedGRU) on the global datasets with 100 epochs. The loss weight is set as {0:1, 1:2, 2:40} for different state (normal 0, wild 1, alert 2). From the results, the loss of NonFed is lower than both FedGRU and FedLSTM, this can be explained by the following reasons: (1) trade-off between communication and accuracy. As the centralized model (NonFed) trained and tested on the global datasets, where it does not require any communication between global and local model. But the FedGRU and FedLSTM need sacrifice certain accuracy to guarantee the efficiency of communication; (2) trade-off between privacy and accuracy. NonFed training already requires tuning parameters like learning rate, momentum, batch size, and regularization. FedGRU and FedLSTM adds potentially more hyperparameters, such as separate tuning of the global model update rule and local SWP optimizer, number of SWP selected per round, number of local steps per round, configuration of update compression algorithms, and more. To preserve the privacy of such hyperparameters may sacrifice the accuracy. FedLSTM 2 actually is a optimal trade-off between accuracy, communication and privacy in this experiment. We may also find that the loss of FedLSTM 2 is lower than FedGRU 2. The reason could be that GRU has fewer parameters and may come at the cost of decreased expressibility. However, the FedLSTM displays much greater volatility throughout its gradient descent compared to the FedGRU model. It can be explained that there are more gates in LSTM for the gradients to flow through, causing steady progress to be more difficult to maintain after many epochs. FedLSTM 2 is not much different from the baseline NonFed LSTM, proving that the FedSWP with LSTM has reasonable convergence and stability.

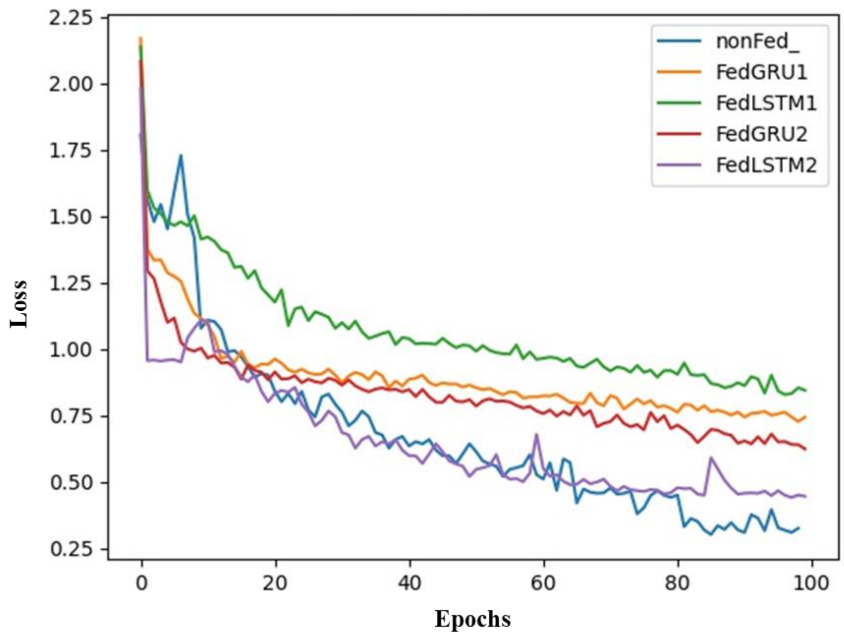

Figure 6. Loss curves for NonFed, FedLSTM, and FedGRU on global datasets

4.3.2 Privacy Analysis

Based on previously clarified privacy in the problem definition section, the privacy-preserving capacities of the proposed FedSWP can be illustrated as follows.

- Data Access: a hybrid model (MTCNN-MobileNet-LSTM) is proposed under the FedSWP framework, demonstrating a distributed privacy-preserving solution for construction OHS management. Notably, this hybrid model achieves accurate facial fatigue prediction by gathering encrypted model parameters instead of directly capturing the original data (private facial image/video), keeping the training dataset locally, and ensuring privacy-preserving of the raw data.

- Model Performance: The evaluation results indicate that the hybrid model's performance under FedSWP outperforms the baseline hybrid model. The baseline model is centralized, requiring aggregate substantial raw data for high-accuracy facial fatigue prediction. Usually, there is a trade-off between the accuracy of facial fatigue prediction and privacy. However, the FedSWP achieves better prediction performance than a centralized approach considering the personalization and privacy-preserving. In general, FedSWP has been proved to achieve accurate prediction of facial fatigue without compromising privacy.

5. Discussion

FedSWP is a general framework for preserving the personal image information of construction workers for OHS management. Compared with the previous works, there are three aspects to the proposed FedSWP's novelty summarized as follows.

- Firstly, FedSWP provides a federated learning-based privacy-preserving solution to incent each construction worker to use IoT devices for monitoring their OHS statuses. Since the OHS-related image data privacy has raised great concerns among construction workers with the expanding adoption of IoT and computer vision technologies in construction sites. Previous studies use the camera to capture unsafe behaviors or adopt the RFID/GPS sensors to capture each construction worker's dangerous positions. These bring the ethics risks of data leakage and can not receive positive cooperations of construction workers to collect more data for a better machine learning model.
- Secondly, FedSWP can offer a more personalized and accurate model by transfer learning to monitor and predict each construction worker's facial fatigue status during their task

execution processes. Current machine learning models for fatigue monitoring mainly train models on all the construction worker's image data and may perform unsatisfied accuracy on a new worker. It may result from the distribution difference of features between the new worker's data and aggregated data.

- Thirdly, FedSWP has tested its accuracy on the task of crane operator fatigue monitoring by using a hybrid model (MTCNN-MobileNet-LSTM). It is one of the most privacy-related and complex tasks as it involves very sensitive facial expressions and very complex spatial-temporal features. The results indicate that FedSWP can work on complex machine learning tasks.

Despite these innovations, our study still has several limitations.

- Firstly, only homomorphic encryption is used in FedSWP for securing the parameter sharing and model distribution. As limited by this hybrid deep neural networks' complexity, in-depth security analysis and more advanced encryption algorithms have not been explored in this study.
- Secondly, FedSWP is only tested in three subjects. In the real situation, hundreds of construction workers' fatigue statuses should be monitored, which may increase the communication cost for FedSWP. Thus, the random sub-sampling mechanism should be designed to improve the efficiency of each round of training.
- Thirdly, as limited by the available dataset for crane operator fatigue monitoring, the dataset used in this study is only collected from drivers. Although features of facial expressions in drivers are similar to crane operators, the real facial image datasets for construction workers are needed in future studies

6. Conclusion

Smark work packaging (SWP) has been proven to be a general model for managing (e.g., model, optimize, and monitor) isolated OHS data and making the OHS insights (e.g., predictions, warnings) ready for each distributed construction worker before task executions at the work package level. However, current machine learning techniques used in SWP for modeling, optimizing, and monitoring need to share or aggregate data from each construction worker. It poses a risk to private data leakage and also can not provide personalized OHS status monitoring.

Thus, this study presents a federated transfer learning framework for SWP, namely FedSWP, to aggregate the encrypted image data parameters from different SWPs of construction workers without compromising privacy and build a personalized model for each construction worker via knowledge transfer. A hybrid deep neural network (MTCNN, MobileNet, LSTM) in FedSWP is adopted to test whether the accuracy of federated transfer learning is better than traditional machine learning methods in facial fatigue monitoring and prediction. Thereinto, five tasks are involved. Firstly, this hybrid deep neural network is trained on public datasets to get a global model. Secondly, this initial global model is spread with homomorphic encryption to each of three selected SWPs. Thirdly, the local model is trained on each SWP's image database. In addition, SWP updates the local models' parameters with homomorphic encryption to create a new global model. Finally, each SWP can get the personalized model by performing transfer learning from the global model to the local model. The evaluation results indicate that the proposed FedSWP outperforms the NonFed hybrid deep neural network and is comparable to GRU with privacy well-preserved.

Future research works as follows are also recommended to enrich the FedSWP.

- To further improve the privacy, security, and trust of FedSWP from construction workers, blockchain technology including cryptography, consensus, and incentive mechanisms (Kim et al., 2019) can be integrated into FedSWP to enhance the security of distributed SWP database and shared parameters.
- To further improve the efficiency and reduce the communication costs of FedSWP, incremental learning could be considered to speed up the model updates. Also, the random sub-sampling mechanism should be designed (Konečný et al., 2016).
- To further validate FedSWP in facial fatigue monitoring for construction workers, the datasets from more construction workers are required to be tested.
- To further generalize FedSWP to other applications of OHS monitoring and prediction. Location data, biosignals, motion data can be tested individually or together as multimodel machine learning to enhance FedSWP.

Acknowledgment

This research was supported by the National Natural Science Foundation of China (NSFC) (Grant No. 71801159 and No. 52078302), the National Natural Science Foundation of Guangdong Province (Grant No. 2018A030310534), Youth Fund of Humanities and Social Sciences Research of the Ministry of Education (Grant No. 18YJCZH090) ,and the funding support from Shenzhen Science and Technology Innovation Commission (Grant No.JCYJ20190808174409266).

Reference

Abtahi, S., Omidyeganeh, M., Shirmohammadi, S., & Hariri, B. (2014, March). YawDD: A yawning detection dataset. In *Proceedings of the 5th ACM Multimedia Systems Conference* (pp. 24-28).

Ahn, C. R., Lee, S., Sun, C., Jebelli, H., Yang, K., & Choi, B. (2019). Wearable sensing technology applications in construction safety and health. *Journal of Construction Engineering and Management*, *145*(11), 03119007.

Aono, Y., Hayashi, T., Wang, L., & Moriai, S. (2017). Privacy-preserving deep learning via additively homomorphic encryption. *IEEE Transactions on Information Forensics and Security*, *13*(5), 1333-1345.

Bickel, S., Brückner, M., & Scheffer, T. (2009). Discriminative learning under covariate shift. *Journal of Machine Learning Research*, *10*(9).

Chang, W. J., Chen, L. B., & Chiou, Y. Z. (2018). Design and implementation of a drowsiness-fatigue-detection system based on wearable smart glasses to increase road safety. *IEEE Transactions on Consumer Electronics*, *64*(4), 461-469.

Chen, Y., Qin, X., Wang, J., Yu, C., & Gao, W. (2020). Fedhealth: A federated transfer learning framework for wearable healthcare. *IEEE Intelligent Systems*.

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*.

Choi, B., Hwang, S., & Lee, S. (2017). What drives construction workers' acceptance of wearable technologies in the workplace?: Indoor localization and wearable health devices for occupational safety and health. *Automation in Construction*, *84*, 31-41.

Ding, L., Fang, W., Luo, H., Love, P. E., Zhong, B., & Ouyang, X. (2018). A deep hybrid learning model to detect unsafe behavior: Integrating convolution neural networks and long short-term memory. *Automation in Construction*, *86*, 118-124.

Edirisinghe, R. (2019). Digital skin of the construction site. *Engineering, Construction, and Architectural Management*.

Fang, W., Ding, L., Luo, H., & Love, P. E. (2018). Falls from heights: A computer vision-based approach for safety harness detection. *Automation in Construction*, *91*, 53-61.

Fang, W., Love, P. E., Luo, H., & Ding, L. (2020). Computer vision for behavior-based safety in construction: A review and future directions. *Advanced Engineering Informatics*, *43*, 100980.

Gal, Y., & Ghahramani, Z. (2016, June). Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *International Conference on Machine Learning* (pp. 1050-1059).

Graves, A., Jaitly, N., & Mohamed, A. R. (2013, December). Hybrid speech recognition with deep bidirectional LSTM. In *2013 IEEE Workshop on Automatic Speech Recognition and Understanding* (pp. 273-278). IEEE.

Guo, Brian HW, Tak Wing Yiu, and Vicente A. González. "Predicting safety behavior in the construction industry: Development and test of an integrative model." *Safety Science* 84 (2016): 1-11.

Hallowell, M. R., Hinze, J. W., Baud, K. C., & Wehle, A. (2013). Proactive construction safety control: Measuring, monitoring, and responding to safety leading indicators. *Journal of Construction Engineering and Management*, *139*(10), 04013010.

Hasanzadeh, S., Esmaeili, B., & Dodd, M. D. (2017). Measuring the impacts of safety knowledge on construction workers' attentional allocation and hazard detection using remote eye-tracking technology. *Journal of Management in Engineering*, *33*(5), 04017024.

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., ... & Adam, H. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*.

Kim, H., Park, J., Bennis, M., & Kim, S. L. (2019). Blockchained on-device federated learning. *IEEE Communications Letters*, *24*(6), 1279-1283.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Konečný, Jakub, H. Brendan McMahan, Felix X. Yu, Peter Richtárik, Ananda Theertha Suresh, and Dave Bacon. "Federated learning: Strategies for improving communication efficiency." *arXiv preprint arXiv:1610.05492* (2016).

Lee, W., Lin, K. Y., Seto, E., & Migliaccio, G. C. (2017). Wearable sensors for monitoring on-duty and off-duty worker physiological status and activities in construction. *Automation in Construction*, *83*, 341-353.

Li, L., Fan, Y., Tse, M., & Lin, K. Y. (2020a). A review of applications in federated learning. *Computers & Industrial Engineering*, 106854.

Li, X., Yi, W., Chi, H. L., Wang, X., & Chan, A. P. (2018). A critical review of virtual and augmented reality (VR/AR) applications in construction safety. *Automation in Construction*, *86*, 150-162.

Li, X., Chi, H. L., Zhang, W., & Shen, G. Q. (2019a). Monitoring and Alerting of Crane Operator Fatigue Using Hybrid Deep Neural Networks in the Prefabricated Products Assembly Process. In *ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction* (Vol. 36, pp. 680-687). IAARC Publications.

Li, X., Shen, G. Q., Wu, P., Xue, F., Chi, H. L., & Li, C. Z. (2019b). Developing a conceptual framework of smart work packaging for constraints management in prefabrication housing production. *Advanced Engineering Informatics*, *42*, 100938.

Li, X., Wu, C., Wu, P., Xiang, L., Shen, G. Q., Vick, S., & Li, C. Z. (2019c). Smart Work Packaging enabled constraints modeling for on-site assembly process of prefabrication housing production. *Journal of Cleaner Production*, *239*, 117991.

Li, X. (2019). Smart work packaging for constraints management in prefabrication housing production (doctoral thesis). The Hong Kong Polytechnic University, Hong Kong, China. https://theses.lib.polyu.edu.hk/handle/200/10191

Li, X., Liu, X., Li, C. Z., Hu, Z., Shen, G. Q., & Huang, Z. (2019d). Foundation pit displacement monitoring and prediction using least squares support vector machines based on multi-point measurement. *Structural Health Monitoring*, *18*(3), 715-724.

Li, X., Chi, H. L., Wu, P., & Shen, G. Q. (2020b). Smart work packaging-enabled constraint-free path re-planning for tower crane in prefabricated products assembly process. *Advanced Engineering Informatics*, *43*, 101008.

Liu, P., Chi, H. L., Li, X., & Li, D. (2020, November). Development of a Fatigue Detection and Early Warning System for Crane Operators: A Preliminary Study. In *Construction Research Congress 2020: Computer Applications* (pp. 106-115). Reston, VA: American Society of Civil Engineers.

Liu, Y., Kang, Y., Xing, C., Chen, T., & Yang, Q. (2020). A Secure Federated Transfer Learning Framework. *IEEE Intelligent Systems*.

Long, M., Wang, J., Ding, G., Sun, J., & Yu, P. S. (2013). Transfer feature learning with joint distribution adaptation. In *Proceedings of the IEEE international conference on computer vision* (pp. 2200-2207).

Lu, Y., Huang, X., Dai, Y., Maharjan, S., & Zhang, Y. (2019). Blockchain and federated learning for privacy-preserved data sharing in industrial IoT. *IEEE Transactions on Industrial Informatics*, *16*(6), 4177-4186.

Luo, X., Li, H., Wang, H., Wu, Z., Dai, F., & Cao, D. (2019). Vision-based detection and visualization of dynamic workspaces. *Automation in Construction*, *104*, 1-13.

Lyu, J., Yuan, Z., & Chen, D. (2018). Long-term multi-granularity deep framework for driver drowsiness detection. *arXiv preprint arXiv:1801.02325*.

Mostafa, K., & Hegazy, T. (2021). Review of image-based analysis and applications in construction. *Automation in Construction*, *122*, 103516.

Niu, Y., Lu, W., Xue, F., Liu, D., Chen, K., Fang, D., & Anumba, C. (2019). Towards the “third wave”: An SCO-enabled occupational health and safety management system for construction. *Safety Science*, *111*, 213-223.

Park, C. S., & Kim, H. J. (2013). A framework for construction safety management and visualization system. *Automation in Construction*, *33*, 95-103.

Sattler, F., Wiedemann, S., Müller, K. R., & Samek, W. (2019). Robust and communication-efficient federated learning from non-iid data. *IEEE transactions on neural networks and learning systems*.

Seo, J., Han, S., Lee, S., & Kim, H. (2015). Computer vision techniques for construction safety and health monitoring. *Advanced Engineering Informatics*, *29*(2), 239-251.

Son, H., Choi, H., Seong, H., & Kim, C. (2019). Detection of construction workers under varying poses and changing background in image sequences via very deep residual networks. *Automation in Construction*, *99*, 27-38.

Sun, B., & Saenko, K. (2016, October). Deep coral: Correlation alignment for deep domain adaptation. In *European conference on computer vision* (pp. 443-450). Springer, Cham.

Tang, S., Roberts, D., & Golparvar-Fard, M. (2020). Human-object interaction recognition for automatic construction site safety inspection. *Automation in Construction*, *120*, 103356.

Tixier, A. J. P., Hallowell, M. R., Rajagopalan, B., & Bowman, D. (2016). Application of machine learning to construction injury prediction. *Automation in construction*, *69*, 102-114.

Voigt, P., & Von dem Bussche, A. (2017). The EU general data protection regulation (GDPR). *A Practical Guide, 1st Ed., Cham: Springer International Publishing*.

Weiss, K., Khoshgoftaar, T. M., & Wang, D. (2016). A survey of transfer learning. *Journal of Big data*, *3*(1), 9.

Wu, J., Cai, N., Chen, W., Wang, H., & Wang, G. (2019). Automatic detection of hardhats worn by construction personnel: A deep learning approach and benchmark dataset. *Automation in Construction*, *106*, 102894.

Yan, X., Li, H., Wang, C., Seo, J., Zhang, H., & Wang, H. (2017). Development of ergonomic posture recognition technique based on 2D ordinary camera for construction hazard prevention through view-invariant features in 2D skeleton motion. *Advanced Engineering Informatics*, *34*, 152-163.

Yang, Q., Liu, Y., Chen, T., & Tong, Y. (2019). Federated machine learning: Concept and applications. *ACM Transactions on Intelligent Systems and Technology (TIST)*, *10*(2), 1-19.

Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. (2014). How transferable are features in deep neural networks?. In *Advances in neural information processing systems* (pp. 3320-3328).

Yu, Y., Li, H., Yang, X., Kong, L., Luo, X., & Wong, A. Y. (2019). An automatic and non-invasive physical fatigue assessment method for construction workers. *Automation in construction*, *103*, 1-12.

Zhang, K., Zhang, Z., Li, Z., & Qiao, Y. (2016). Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Processing Letters*, *23*(10), 1499-1503.

Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., ... & He, Q. (2020). A comprehensive survey on transfer learning. *Proceedings of the IEEE*.