Federated Smart Work Package Framework with Triplet Loss for Mental Fatigue Monitoring

Jianhuan Zeng[1], Xiao Li[*2] M. ASCE, Maxwell Fordjour Antwi-Afari[3]

ABSTRACT

As modular construction projects and prefabrication have become increasingly prevalent, safety in heavy crane operation during on-site assembly has become a pressing concern. A significant safety concern is monitoring the mental fatigue of crane operators. Previous research has proposed a federated transfer learning-enabled smart work packaging (FedSWP) to achieve personalized and privacy-preserving fatigue monitoring. However, when applying FedSWP for fatigue monitoring, network efficiency issues are magnified in data-insufficient and resource-constrained smart work packages (SWPs). This paper introduces an adaptive and lightweight federated smart work package framework with triplet loss (FedSWP-TL), explicitly focusing on efficiently recognizing crane operators' mental fatigue. By leveraging triplet networks and innovative techniques such as compressive aggregation and a lighter MobileVit network, the FedSWP-TL demonstrates enhanced generalization and mobility capabilities. Through evaluation of YAWDD, DROZY, and ConPPMF datasets, our approach outperforms efficiency while monitoring fatigue status under constrained resources. The results highlight the adaptability of FedSWP-TL to diverse groups of SWPs, showcasing its potential for practical implementation in data-insufficient environments.

[1] PhD Student, Department of Civil Engineering, The University of Hong Kong, Hong Kong SAR, China.

[2] Assistant Professor, Department of Civil Engineering, The University of Hong Kong, Hong Kong SAR, China. Email: shell.x.li@hku.hk

[3] Lecturer, Department of Civil Engineering, Aston University, Birmingham, United Kingdom

Introduction

As cranes are one of the most demanded and utilized pieces of equipment in construction projects worldwide, the safety of crane operation remains a critical concern due to the demanding nature and the potential hazards. Crane operators' fatigue should be closely monitored, especially since their work requires constant concentration and alertness (Li et al., 2019; Liu et al., 2021). Previous studies have developed various deep learning (DL) models for fatigue prediction (Imran et al., 2024). Meanwhile, privacy and data security have become paramount as fatigue detection on DL models mainly involves using private data like face (Li et al.; Sun et al., 2023), bio-signals (Wang et al., 2023; Guarda et al., 2022), and speech (Yan et al.; Shen et al., 2021). Many regions and countries are enacting strict laws, policies, and regulations on privacy and data security (Li et al., 2022), such as the European Union's General Data Protection Regulation (GDPR), China's Personal Information Protection Law, and the California Consumer Privacy Act. These regulations impede DL training with isolated data repositories in many fields, including construction. Thus, there is a pressing need for crane operators' fatigue detection in a personalized and privacy-preserving way.

Federated learning (FL) is a distributed machine learning that encourages using Internet of Things (IoT) devices for safety monitoring while preserving the privacy of sensitive information. For example, Li et al. (2021) introduced a federated transfer learning-enabled smart work package framework (FedSWP). This distributed vision-based approach leverages personal images of individuals' faces to detect signs of mental fatigue while safeguarding their confidentiality. A SWP is the smallest distributed agent facilitating task planning, scheduling, and execution (Li et al., 2020) via advanced artificial intelligence (AI). It consists of a cluster of intelligent resources such as non-intrusive wearable bio-sensors, face cameras, limited bandwidth, constrained storage, AI models for fatigue detection, and a small private dataset. SWP can then transmit personal data to the cloud for training and offer safety alerts or health

insights. The FedSWP framework employs transfer learning with a hybrid model to create personalized models. There is other FL methods applied to areas such as building energy consumption (Li et al. 2022), human-robot collaborative assembly tasks (Cai et al. 2024), and quality defect inspection (Wu et al. 2024) to protect privacy in construction. Existing methods have relatively underestimated the need for an adaptive and lighter framework for resource-constrained devices and various construction projects, while the federated nature of learning has further amplified efficiency challenges with limited local data. Moreover, triplet loss for creating robust fatigue embedding representations, especially under federated frameworks, remains unexplored to the best of the researchers' knowledge. Under a non-central learning paradigm, finding a more efficient fatigue detection solution seamlessly adaptive to diverse construction projects is crucial among data-insufficient and resource-constrained SWPs.

Two key issues hinder the feasibility of mental fatigue monitoring with federated and personalized AI models. Firstly, resource-constrained hardware faces the challenge of balancing higher accuracy with smaller computation and communication. For example, hybrid deep neural networks with long short-term memory (LSTM) for temporal features can result in longer times and increased energy consumption, which may overwhelm the hardware of SWPs (Chou et al., 2025). Secondly, data-insufficient SWPs aggravate the gaps in adaptation to new groups. SWPs may encompass videos of various appearances in diverse backgrounds, different cameras, and heterogeneous computing chips, introducing biases during non-independent and identically distributed (non-IID) training (Xu et al., 2023). This ultimately leads to a degradation of accuracy in new SWPs or an overfitting issue after personalized training in small local datasets. While federated learning provides a privacy-enhancing solution for monitoring fatigue, it does not effectively adapt to different groups of SWPs because of capability and statistical heterogeneity. Given the limited resources and data scarcity of each remote SWP with varying characteristics (e.g., various sensors), finding an alternative solution with adaptive and

mobile attributes is imperative in a model application for fatigue monitoring systems to crane operators.

This study aims to introduce a federated smart work package framework with triplet loss (FedSWP-TL) to facilitate mobility and generality during on-site mental fatigue monitoring of crane operators. To this end, the research seeks to (1) utilize a triplet network for each SWP to learn high-order embedding representations for fatigue, (2) provide a mobile-friendly federated transfer learning framework with compressive communication and efficient computation, (3) validate the generality and mobility of FedSWP-TL in the context of fatigue monitoring on datasets of YawDD, DROZY, and newly established ConPPMF. The remainder of this study is structured as follows. Section 2 provides a comprehensive overview of past studies. Section 3 introduces the newly proposed FedSWP-TL framework. Section 4 presents the experiment. Results for generality and mobility are evaluated in section 5, while section 6 highlights the contributions of the research and discusses its limitations and potential application. Finally, the conclusion is summarized in Section 7.

Related Works

This section reviews signal types for mental fatigue monitoring, explores detection methods, and briefly overviews secure federated communication for efficiency and deep metric learning for adaptability. Research gaps are summarized at the end.

Mental Fatigue Signals

Mental fatigue refers to the depletion of mental alertness, demonstrated by the incapacity of reaction time, coordination, judgment, or concentration (Li et al., 2019). Various signals for objective and practical assessment include physiological indicators, operational behaviors, and visual characteristics. Physiological indicators include but are not limited to heart rate variability (HRV), respiration rate (RR), electrocardiograms (ECG), electromyograms (EMG),

electrooculograms (EOG), and electroencephalograms (EEG). EEG is widely acknowledged as a dependable and crucial assessment tool due to its exceptional temporal resolution. For instance, Mehmood et al. (2023) proposed a wearable EEG-based device to quantitatively assess workers' mental fatigue levels. Additionally, Wang et al. (2023) utilized innovative cushions to extract the HRV and RR features to model the mental fatigue status of the participants. These physiological indicators are valuable but require non-intrusive sensors and robust data collection for real-time monitoring (Imran, 2024).

Besides, fatigue could be inferred from operational behaviors. Elevated risk of errors and slower reaction times are common indicators of mental fatigue, leading to decreased operational performance (Xu et al., 2023). Even though analyzing operational behavior can occasionally provide insights into fatigue states (Liu et al., 2021), the real-time collection and evaluation of operational behaviors remain challenging due to the diversity of workers' roles and tasks.

In addition to physiological and operational signals, fatigue monitoring can rely on visual characteristics as indirect physical measures. These non-invasive approaches offer a more practical means of real-time monitoring and can be easily integrated into daily activities. For instance, head nodding, eye tracking, pupil size changes, and jaw state can be monitored using cameras or facial recognition technology (Liu et al.; Dziuda et al., 2021; Li et al.; Lu et al.; Sun et al., 2023). Although not as accurate as biometric EEG methods, these indirect measures deliver insightful fatigue information. Those visual characteristics can be combined with physiological or operational signals to enhance the overall fatigue assessment.

Methods for Fatigue Detection

Traditionally, the assessment of mental fatigue has relied on self-reporting measures such as Chalder Fatigue Scale (Chalder et al.,1993), Fatigue Assessment Scale (FAS, Michielsen et al.,2003), NASA-TLX (Hart, 2006), and Karolinska Sleepiness Scale (KSS, Kaida et al., 2006).

While these approaches are straightforward, they lack precision due to potential biases in subjective reporting. Moreover, they are unsuitable for continuous monitoring as workers must periodically complete fatigue-related questionnaires.

With the advancement of DL, indirect measures have embraced impartial and non-intrusive technology for analyzing mental fatigue signals. The cognitive fatigue status can be aptly determined with greater precision as the computer autonomously learns key features such as facial expressions, head movements, eye states, physiological patterns, and working performance (Lu et al., 2023; Yang et al., 2024).

DL has achieved remarkable success in visual classification tasks, particularly with convolutional neural networks (CNN), such as facial feature fusion convolutional neural network for driver fatigue detection (Sun et al., 2023) and CNN-Attention Structure using EEG data (Xu et al., 2021). However, capturing temporal features for fatigue detection is necessary, as it involves a continuous process rather than a one-time event. LSTM or gated recurrent unit (GRU) architectures are incorporated (Lu et al., 2023). In the federated transfer learning framework, FedSWP uses a hybrid model architecture with a face detector (MTCNN) and spatial and temporal feature extractors (MobileNet, LSTM). Although MTCNN and MobileNet are efficient, the computational demand increases for the complexity of integration and additional temporal LSTM, particularly in resource-constrained environments. Vision transformers (ViTs, Dosovitskiy et al., 2021) are a more sophisticated option for learning sequential visual representations. Even though these architectures are known for their high accuracy, they are cumbrous (e.g., ViT-B/16 having 7.5 million parameters) and difficult to optimize when working with real-world data (Mehta& Rastegari, 2021; Guo et al., 2022). Lightweight studies are newly developed to learn spatial and temporal features on mobile platforms, such as MobileFormer and MobileViT. MobileViT focuses on minimizing parameter

count, while MobileFormer optimizes for FLOPs. However, little research has been conducted on mobile models capturing spatial and temporal features for fatigue monitoring.

Secure Federated Communication

Despite methods categorized by model architecture, typical intelligent monitoring methods can be divided by learning paradigm, i.e., centralized, federated, and decentralized. A federated learning solution for privacy-preserving fatigue monitoring guarantees participant privacy by sharing only gradients across devices, while sensitive training data remains on the local device without any raw features or labels being communicated between parties (Li et al., 2021). Three main defense approaches to prevent reconstruction attacks during parameter sharing are cryptography-based, infrastructure-based, and perturbation methods. Cryptography-based methods, such as secure multi-party computation and homomorphic encryption (Moriai, 2019), pose a computational challenge and may hinder participant communication. Infrastructure-based methods, such as trusted execution environments (TEE) (Subramanian et al., 2017), require specific hardware and may not be practical in all scenarios. Perturbation techniques involve random perturbation, message shuffling, compression, and sparsification (Yang et al., 2023). Compression has been shown to compress gradients over 300 times without sacrificing accuracy (Zhu et al., 2020).

FedSWP (Li et al., 2021) utilizes homomorphic encryption to merge model parameters, further amplifying the bandwidth requirements since communication among SWPs can be expensive. When multiple SWPs need to interact with the server concurrently, they may be queued to receive their updates. Even though transfer learning is applied to minimize communication, the size of the encrypted messages can be substantially larger than their plaintext counterparts, exacerbating the communication overhead and potentially leading to network congestion. Therefore, resource-constrained SWPs require a more efficient communication strategy.

A decentralized learning solution for privacy preservation, like decentralized adaptive work package (DAWP) learning for construction occupational health and safety (COHS) monitoring (Li et al. 2024), ensures participant privacy through local datasets and decentralized model updates, free from the limited bandwidth of the centralized aggregator in FL solutions. DAWP leverages blockchain technology to consolidate model parameters, but the bandwidth requirements still necessitate costly communication between SWPs and blockchain nodes.

Deep Metric Learning

Traditional DL approaches in FedSWP and other mentioned methods for fatigue detection primarily focus on discriminative classification, which is prone to over-fitting, especially when adapting to heterogeneous small datasets. Deep metric learning approaches seek to gain insights into the underlying data patterns by constructing a non-linear embedding space. In this space, the classification of data points into specific classes is determined by evaluating the distances between them. To facilitate this process, prior knowledge has been incorporated into a projection function, which converts the unprocessed input data into a suitable representation for comparison of similarities (Li et al., 2023). Popular deep metric learning networks include pairwise, matching, prototypical, relation, and graph-based networks, employing a feed-forward mechanism to compare similarity.

Pairwise networks are a specific kind of deep metric learning model that is created to process pairs of examples and learn a shared feature space that can be used to distinguish between two classes. For instance, Siamese networks rely on two identical neural networks to learn embeddings from a pair of samples and then calculate a weighted metric to determine their similarity. The Siamese network (Zhu et al., 2022) leverages weight parameter sharing to map features to a shared feature space, enabling effective comparison and analysis of samples from different sources. Triplet Network (Ma et al., 2022) extends Siamese networks to three with

shared parameters to output the comparison probability. By utilizing a distance-based measure and corresponding loss function, those networks can cluster similar samples together and keep dissimilar samples separated. This enables downstream tasks to be performed more accurately and generally, enhancing the network's overall performance.

Euclidean distance, cosine distance, contrastive loss, and triplet loss commonly measure similarities among pairs or triplet samples. The contrastive loss encourages two inputs belonging to the same class to be closer to each other, while two images with different labels are kept far apart (Ma et al., 2022). The triplet loss is designed to learn a similarity metric or embedding space where similar samples are closer and dissimilar samples are farther apart. There is limited research on deep metric learning approaches to extract fatigue features, promoting resilient and adaptive monitoring in data-insufficient settings.

Research Gaps

Although a non-central learning paradigm for fatigue monitoring is significant, a mobile-friendly framework has not been fully explored. Moreover, little research has been conducted on triplet loss in fatigue monitoring, especially under the adaptation to new groups of operators. Thus, the main research question is how to improve the efficiency of federated solutions in data-insufficient and resource-constrained settings.

Methods

The primary objective of this study is to determine the capability of triplet loss to differentiate non-fatigued and fatigued states for unseen operators in heterogeneous nodes. The secondary objective is to design a mobile-friendly federated transfer learning framework with compressive communication and efficient computation. To facilitate generality and mobility, this section presents a federated smart work package framework with triplet loss that incorporates compressive aggregation and MobileVit triplet network architecture.

This section delineates the methodologies employed in crafting the FedSWP-TL, encompassing federated transfer learning within the FedSWP-TL network, the incorporation of the triplet network, MobileVit, and selective gradient aggregation techniques throughout the off-site training phase, as well as the similarity analysis during on-site application for fatigue monitoring.

The FedSWP-TL Network

The FedSWP-TL network is presented in Fig. 1. It comprises a DL node and multiple SWP nodes, each functioning as a computing node with a local dataset. The DL node generally works alongside SWP nodes for monitoring, controlling, and managing. It selects the appropriate SWPs, aggregates model parameters, and serves as the command interface. The SWP operates as an independent contractor, with a team of workers and monitoring videos specific to their tasks. The SWP node trains a personalized model with its private dataset and the global model.

The federated transfer learning interaction in the FedSWP-TL network: (1) SWP0 with a public dataset D_0 initially shares global weight Θ_{meta}. (2) Then the global weight Θ_{meta} is updated from SWPs with $\mathrm{D} = \{D_k\}_k^{\mathrm{K}}$. This collaborative effort among K SWP continues until the DL node has successfully fine-tuned the model to an acceptable level. (3) A new SWP could join the FedSWP-TL network, interacting with the meta-model.

Off-site Training Process

The FedSWP-TL off-site training process (Fig. 2) follows the general FedSWP process. It trains a global model on a public dataset and then transfers the initial model to local SWP nodes. Each SWP node then trains a local model on its database, and SWPs send local models' parameters to update the global model. Ultimately, personalized models are independently obtained. However, the FedSWP-TL is more advanced with triplet loss, compressive aggregation, and MobileVit architecture.

As shown in Algorithm 1, before the training process, triplets $\{T = (x, x_+, x_-) | \, i \in \mathrm{N}\}$ are created as input on line 1. Remarkably, the global weight Θ_{meta} is more portable than that of a hybrid model with LSTM on line 4. Θ_{meta} in the triplet network is updated via a compressive aggregation schema instead of an encrypted federated averaging algorithm on lines 6-9 and 17-18. The loss is not only classification loss but the sum of classification loss and triplet loss on line 15. The enhanced version of FedSWP with Triplet loss is lightweight and adaptable to monitor fatigue without compromising data privacy and model personalization.

The adaptive SWP: The key difference between our approach and traditional ones is to learn individual fatigue embeddings via a triplet loss. The triplet loss is to train encoders to highlight the most discriminative representation of fatigue. Fatigued and regular faces may look similar (like the same person) but pose contrasting fatigue properties. The global encoder can prioritize the fatigue expressions over appearances, and the personalized SWP could capture the individual differences in fatigue response patterns. A new SWP in different construction projects and sites could use the global encoder and personalize its model with a limited training dataset due to common fatigue responses in humans.

Let $(\mathrm{x_j}, \mathrm{y_j})$ be the j-th sample in a training set $\mathrm{D_k}$. During the neural network training, training samples are selected and formed into triplets $\{T_i = (x, x_+, x_-) | \, i \in \mathrm{N}\}$ with an anchor sample x, a positive sample x_+(similar to the anchor) and a negative sample x_-(dissimilar to the anchor) where the relative label satisfies $y = y_+ \neq y_-$.The predicted label $f(x) = f(x | \, \Theta_k)$. The model parameters are a combination of global and personalized parameters, $\Theta_k := \{\Theta_{\mathrm{meta}}, \boldsymbol{\theta}_{L-1}, \boldsymbol{\theta}_L\}$, where Θ_{meta} represents the global weight, $\boldsymbol{\theta}_{L-1}, \boldsymbol{\theta}_L$ are personalized

weights on the last dense layers. The model parameter vector $\Theta_k \in \mathbb{R}^{\mathrm{d}}$ (e.g., weight and bias) can be obtained by minimizing the loss function (Eq. 1):

$$\mathcal{L} = \mathcal{L}_T\,(T_i) + \mathcal{L}(f(x), y) \tag{1}$$

The loss function for the model is a combination of two losses: The triplet loss $\mathcal{L}_T\,(T_i)$ for the embedding purpose and binary cross entropy loss $\mathcal{L}(f(x), y)$ for the prediction purpose. $\mathcal{L}_T\,(T_i)$ is a function (Eq. 2) used in machine learning, particularly in face recognition, image retrieval, and similarity learning tasks. $\mathcal{L}(f(x), y)$ is carried out to enhance the alignment between the predicted labels and the ground truth in the classification process.

$$\mathcal{L}_T\,(T_i) = \mathrm{Max}[\mathrm{d}(x, x_+) - \mathrm{d}(x, x_-) + \mathrm{m}, 0] \tag{2}$$

where $\mathrm{d}(\cdot)$ represents a distance metric (e.g., Euclidean distance or cosine similarity), the margin m is a hyperparameter that defines the minimum desired separation between the positive and negative samples.

The lightweight SWP: The lightweight attribute is ameliorated by deploying a mobile model architecture and compressive aggregation to minimize computation and communication. Due to its proven superiority (Mehta & Rastegari, 2022; Li et al., 2024), this study utilizes MobileVit as the backbone to help with resource-constrained training. MobileViT combines the strengths of CNNs and ViTs to build a lightweight network, while low FLOPs in MobileFormer do not necessarily result in low latency (Vasu et al., 2023). MobileVit consists of two convolution layers, six MobileNetv2 (MV2) blocks, and three MobileViT blocks (See Table 1). A global pooling layer then completes the model to indicate the feature embeddings of each input video. Notably, a fully connected layer in the original setting is replaced with two personalized dense layers, where SWPs will hold independently. The MobileViT block will capture both local and global information by utilizing a conventional convolutional layer followed by a point-wise

convolutional layer. MV2 blocks are mainly deployed for down-sampling, ↓2 means that down-sampling to the half.

As shown in lines 17-18 in Algorithm 1, the lightweight attribute is also augmented by minimizing the communication costs during the update of global weights. The gradients computed by each SWP were communicated to a DL node for merging and updating. This can be a bottleneck during training, especially in SWP scenarios with limited network bandwidth and constrained Input/Output(I/O) resources. FedSWP-TL addresses this issue by compressing the gradients before transmission. It utilizes sparsification and quantization to reduce the size of the gradient tensors. Sparsification involves identifying and transmitting only a subset of the gradients with significant magnitudes, while quantization reduces the precision of the gradient values (Yang et al., 2023). In addition to less communication, the transmitted gradient is harder for attackers to leverage to reconstruct the raw image or infer sensitive information.

On-site Monitoring Process

When transitioning the FedSWP-TL framework from experimental presentation to actual application, each SWP possesses an individualized trained model, with continuous facial recordings being categorized using classification prediction and nearest-neighbor analysis. The identification of fatigue is further interpreted by calculating similarities between the present embedding and the average embeddings for fatigue and non-fatigue.

During on-site monitoring, two kinds of SWPs are observed: existing SWPs with integrated trained models and new SWPs that initially retrieve the global model from the DL node and then fine-tune personalized layers. The network demand for newly joined SWPs is initially high due to the global model deployment, but it could diminish as the personalized updates become internalized.

Experiments

Models of FedSWP-TL consist of a global model in SWP0 and personalized models in the other SWPs, tested across private datasets. The global model is initially trained on the public dataset, then is fine-tuned and tested on private datasets, while the personalized models are fine-tuned and tested on private datasets. The origin FedSWP serves as a benchmark method to showcase the enhanced adaptive and lightweight characteristics.

Datasets

Multiple datasets naturally represent realistic and heterogeneous local datasets in different SWPs, enabling a thorough exploration of FedSWP-TL performance across diverse sensors and heterogeneous adaptions. Each dataset used in the study captures a different scenario related to fatigue monitoring. YawDD provides instances of yawning under controlled conditions, DROZY offers diverse and challenging simulated scenarios, and ConPPMF presents real-world data from construction environments. Table 2 summarizes the various settings of the three databases, with the *Age* column reflecting the age range of the majority of the subjects.

YawDD, short for "Yawing Detection Dataset," comprises two datasets featuring videos captured at a resolution of 640 × 480 pixels with 24-bit actual color (RGB) from car operations in 2014. It is identical to alerting fatigue of the crane operator and supports at least three papers (Li et al., 2021, 2022; Liu et al.,2021). These videos depict a range of facial expressions, including regular, talking/singing, and yawning states at 30 frames per second (FPS). The first dataset, obtained from the front mirror perspective, comprises 270 videos featuring 90 subjects (47 males and 43 females). The second dataset, captured from the dashboard viewpoint, involves 29 subjects (16 males and 13 females).

DROZY, as the abbreviation of "The ULg Multimodality Drowsiness Database" (Massoz et al., 2016), stands out as the most widely utilized dataset for monitoring mental fatigue. It is also

valuable for monitoring the mental fatigue of crane operators. The dataset is meticulously gathered through a comprehensive approach involving the administration of the psychomotor vigilance test (PVT), the application of polysomnography (PSG) electrodes, and the assessment of sleepiness levels using the KSS in a controlled laboratory setting over two days. The study cohort included 14 healthy participants, consisting of 3 males and 11 females, with an average age of 22.7 years (±2.3 SD). The dataset encompasses multiple modalities, featuring EEG data from 5 channels (Fz, Pz, Cz, C3, and C4), other physiological signals such as EOG, ECG, and EMG, and camera videos and reaction time evaluations.

ConPPMF, as the abbreviation of "Construction Datasets for privacy-preserving fatigue monitoring," stands out as the most related dataset for monitoring mental fatigue. The dataset is gathered through a non-invasive approach involving RGB cameras positioned in front of faces and smartwatches worn on wrists to capture visual and physiological data. Three construction workers, an average of 42.7 years old, were recorded during daily operations. Fatigue status is binary labeled based on self-reporting and job performance after recording. Fig. 3 shows some frame samples of monitoring videos in the dataset.

To thoroughly evaluate the performance of FedSWP-TL, 97 subjects were selected from these diverse datasets to form the experimental groups. The videos were reorganized into seven distinct datasets according to the subjects to facilitate intricate situations: a public dataset and six private datasets. A public dataset D_0 includes videos from 66 subjects randomly selected from YawDD datasets to train initial global weight while six private sets $\mathrm{D} = \{D_k\}_k^{\mathrm{K}=6}$serve as local datasets of SWPs. These include four datasets from separate YawDD subjects, one from DROZY subjects, and one from ConPPMF subjects, distributed in different SWP nodes (named SWP1, SWP2 and SWP3, P4, DROZY, and ConPPMF).

Evaluation

To evaluate the FedSWP-TL framework's generalization ability, the performance of personalized models in P4, YawDD, and DROZY nodes are individually tested on local testing samples to simulate real-world adaption to heterogeneous groups of SWPs. The framework's adaptability was assessed by observing those model performances across three distinct databases. Model performance was measured using Recall and F1-score as metrics. Recall, emphasizing the sensitivity of fatigue detection, is crucial for safety monitoring as it focuses on accurately identifying fatigued workers to prevent potential accidents. The wrong judgment increases the workload of managers checking, but the missing detection would lead to undesirable accidents in construction. F1-score, a combination of recall and precision, provides a balanced view of detection sensitivity and accuracy in safety monitoring applications.

In addition to generalization, the study also examined the mobility ability of the FedSWP-TL framework. This study analyzes the balance between performance and resource requirements to assess the framework's ability to maintain promising performance with minimal overhead. Resource evaluation included considerations such as network parameter count, convergence speed, and computing times. The network parameter count reflects the model's memory utilization and portability, with fewer parameters indicating a more efficient and portable model. Convergence speed, representing the time required for model training, influences the speed of model development and deployment. Performance evaluation encompassed key statistical indicators such as accuracy (ACC), the area under the receiver operating characteristic curve (AUC), F1-score, and recall, providing a comprehensive assessment of the FedSWP-TL framework's effectiveness in mental fatigue monitoring. Additional metrics beyond ACC and AUC were included to offer a more nuanced evaluation of the framework's performance in real-world scenarios.

Label harmonization

This paper simplifies the fatigue state into binary classification, making the model easier to train and interpret, especially for initial development stages, requiring fewer computational resources in resource-constrained environments and leading to more robust models in diverse conditions. The existing labels across different datasets are inconsistent and do not align with the triplet's transformation. To accurately represent the transitional states between alertness and fatigue, these datasets are relabeled into two fatigue levels: regular (labeled as 0) and fatigue (labeled as 1). For instance, in the YawDD dataset, behaviors such as stillness, moving head, normal talking, laughing, and singing, which are least associated with fatigue, are relabeled as 0. Original yawning markers with values 1, 2, 3, 4, and 5, indicating varying fatigue levels, are relabeled as 1. The DROZY dataset is labeled by KSS. The labels in ConPPMF videos are binary and checked every 10 frames. The label unification aims to create a consistent framework for the global model to learn from multiple sources, standardizing the data and thereby improving its generalizability across different scenarios. However, this process can lead to a loss of detail, as the manifestations and judgment criteria for fatigue vary between datasets. This inconsistency can result in the global model exhibiting different levels of accuracy and generalization when applied to different datasets, which is then eliminated during personalized fine-tuning for annotated datasets. The accuracy and generalization are enhanced when analyzing results for non-annotated datasets.

Implementation

Frame selection: The videos in these datasets contain a range of frames. Rather than analyzing every frame, this technique selects several fixed, continuous frames that clearly exhibit fatigue or non-fatigue status. This approach reduces computational costs while preserving essential information.

Data splitting: The private frames are divided into 80% training (fine-tuning) samples and 20% testing samples through random sampling, ensuring that both sets maintain an identical distribution of fatigue and non-fatigue samples.

Triplet formation: The training samples are then organized into triplets $\{T_i = (x, x_+, x_-) | \, i \in N\}$. There are two types of triplet samples such as one in the form of fatigue-triplets (fatigue-labeled frames, fatigue-labeled frames, non-fatigue-labeled frames) and the other in the form of non-fatigue-triplets (non-fatigue-labeled frames, non-fatigue-labeled frames, fatigue-labeled frames).

Procedure: The training phase involves using the public dataset to train a robust initial global model capable of detecting fatigue at SWP0 and fine-tuning the global model and personalized models on private datasets among the FedSWP-TL framework. Although artifacts may persist in the input from external sources, rare data-cleaning methods are employed to remove undesired fluctuations, such as sensors on subject faces from DROZY, glasses, or mustache from YAWDD. These artifacts can be mitigated through triplet loss, which helps the model differentiate between fatigue-relevant and irrelevant features. Table 3 displays the statistical details of the processed data.

Hyper-parameters: The SWP models are executed in PyTorch (2.2.1+cu121) using two NVIDIA 4090 24GB GPUs among three 8-core 16G machines. As shown in Table 4, the network optimization uses the Adam optimizer at the initial learning rate of 0.001. The learning rate is decayed by a factor of 0.1 after every five epochs. The batch size is set to 256, and the maximum number of epochs is set to 25.

Results

The study sought to compare the effectiveness of our upgraded FedSWP with the Triplet loss approach against the original FedSWP learning method in fatigue monitoring.

Generalization Ability Evaluation

As demonstrated in Fig. 4, the FedSWP-TL models outperform the baseline on recall performance across all new SWPs (P4, DROZY, and ConPPMF) while maintaining a comparable F1 score. The results show FedSWP-TL's enhanced adaptability to new SWP nodes, encompassing unseen subjects from the global training. Table 5 shows data statistics of private sets in three SWPs.

Both FedSWP-TL and FedSWP methods demonstrated good transferability to the P4 node as F1 scores and recalls were greater than 0.85. Even though the DROZY's mean F1 score of FedSWP-TL was smaller than that of FedSWP, the recall was apparently better than that of FedSWP with a larger value and smaller dispersion. Similarly, the performance in ConPPMF outperforms FedSWP with a comparable F1 score, larger recall, and smaller deviation. The results on DROZY and ConPPMF emphasize the improved adaptability of FedSWP-TL.

Mobility Ability Evaluation

Fig. 5 illustrates a promising reduction in FedSWP-TL overhead compared to traditional FedSWP. FedSWP-TL has 2.32M parameters and converges at around 35 epochs, whereas FedSWP requires 24.6M parameters at a slower convergence speed. Both offline and online training times of FedSWP-TL were approximately two times faster than the baseline. The results with significantly fewer parameters, faster convergence, and reduced training time underscore that FedSWP-TL is a mobile-friendly federated transfer learning framework.

Fig. 6 indicates that the advanced FedSWP-TL approach with a smaller overhead achieved performance comparable to that of the FedSWP method. The performance is estimated in

accuracy, F1 score, AUC, and recall, while overhead is analyzed by the number of parameters and training time. As shown in dashed lines, FedSWP-TL highlighted an approximately 10 times smaller number of parameters and 2 times less training time than FedSWP, highlighting the reduced consumption of the FedSWP-TL framework. When models are trained, the dataset YawDD, FedSWP-TL showed slightly lower accuracy, F1 score, and AUC but better recall than FedSWP. This trend was also observed in the dataset DROZY, where there was lower accuracy and F1 score but better recall. The drop in the F1 score (5% in DROZY and 10% in YawDD) results from lower precision on larger samples and higher recall on small but important samples. Furthermore, FedSWP-TL exhibited higher sensitivity to fatigue detection with a higher recall (+3% roundly) while maintaining comparable accuracy and AUC score. Additionally, the consistent observation of a similar F1 score and better recall of P4 models (Fig. 4) also supports that FedSWP-TL is a competent counterpart to FedSWP. The competent performance under reduced overhead validates the higher efficiency of the FedSWP-TL.

Fairness across SWP nodes

The evaluation of FedSWP-TL considers fairness as an essential criterion. This criterion determines whether FedSWP-TL learning exhibits notable variations on each SWP node for fatigue monitoring tasks. The homogeneity across nodes is shown in Table 6 and Fig. 7, where SWPs present similar prediction performance and similar training convergence using triplet loss in the FL framework.

The similar accuracy of AUC and recall in Table 6 highlights the acceptable heterogeneity of SWPs even when the F1 score is diverged. Moreover, the FedSWP-TL learning fosters fairness in personalization performance across diverse nodes, even on the worst-performing nodes, which is a crucial advantage. Fig. 7 demonstrates that the training of all SWPs in FedSWP-TL performs consistently in terms of convergence. SWP2 demonstrates fast convergence and excellent personalization, achieving a converged AUC of about 0.9 after epoch 15, while other

SWPs have similar convergence speeds (within three epochs). This reveals that FedSWP-TL learning promotes fairness, showing equal distribution of learning progress and performance outcomes across all SWP nodes. Overall, FedSWP-TL learning improves the reliability and resilience of the models across the entire network, contributing to a more equitable and inclusive learning collaboration.

Discussion

Main Contributions

The utilization of non-central learning for crane operators' fatigue monitoring (Li et al. 2021), building energy consumption (Li et al. 2022), human-robot collaborative assembly tasks (Cai et al. 2024), quality defect inspection (Wu et al. 2024) and construction occupational health and safety monitoring (Li et al. 2024) has been proven beneficial in providing robust privacy protection on construction, as demonstrated in previous studies. This paper further enhances generalization and mobility abilities. The FedSWP-TL introduces two innovative facets herein. First, it employs high-order embedding representations derived from triplet loss, enhancing the discriminative performance of fatigue features across various SWPs with limited datasets. Secondly, the framework incorporates a selective gradient compression and a lightweight architecture design, optimizing its suitability for resource-constrained work package-based monitoring while maintaining high personalization performance. By integrating the MobileVit block and employing gradient compression techniques, the model achieves superior performance with reduced computational costs and communication overhead.

Practical Applications

The FedSWP-TL models have demonstrated superior recall performance across all new SWPs while maintaining comparable F1 scores (Fig. 4). This finding confirms that the use of triplet loss facilitates the learning of more effective fatigue representation. The distributions of both methods exhibit less concentricity compared to those on P4, indicating that DROZY and

ConPPMF, unlike P4, are notably distinct from the training group. FedSWP-TL models exhibit improved adaptability to more heterogeneous DROZY and ConPPMF, showcasing better performance (higher recall and F1-score) and more concentration (generally smaller deviation except one in ConPPFM's F1-score). Meanwhile, the baseline models miss more than 25% fatigue samples (with recall values around 0.75) in DROZY and ConPPMF, in contrast to their satisfactory performance in P4 (0.88). The FedSWP results suggest potential unreliability in monitoring distinct SWPs that observe new subjects using diverse sensors. Additionally, it indicates that triplet loss, unlike classification loss, prioritizes fatigue-specific features over general attributes like resolution, background, and identity. Moreover, the fewer parameter counts and faster online training time depicted in Fig. 5 reflect that FedSWP-TL enables swift deployment and adaptability. Consequently, the FedSWP-TL enhances the feasibility of efficient mental fatigue monitoring across varying fatigue patterns and dissimilar SWPs.

Fatigue is typically a continuous state with varying degrees of severity in practical applications. The scorecard method used in financial credit classifications can be adapted for fatigue monitoring to provide a straightforward and interpretable way to convert fatigue probabilities into risk scores. By incorporating the scorecard approach (Eq. 3), the FedSWP-TL moves beyond binary classification to create a more nuanced evaluation for monitoring fatigue states.

$$\text{score} = \text{base score} + \text{scaling factor} * \log(p_+/p_-) \quad (3)$$

Adjust the base score and scaling factor accordingly to the desired scale for fatigue monitoring, which warrants further research in future studies. For instance, if the base score for a non-fatigue state is 650, and the fatigue probability is twice that of the normal probability, the fatigue risk score increases by 50, which is Eq. 4.

$$\text{the risk score} = 650 + 50 * \log(p_{fatigue}/p_{normal}) \quad (4)$$

Personalization Performance

The performance of P4's personalized model (recall/F1-score: 0.95/0.85 as shown in Fig. 4) surpassed that of the global YawDD model (0.95/0.77 as depicted in Fig. 5). This outcome can be attributed to the personalized training conducted on a private set within P4. Interestingly, the recall of DROZY's personalized model (0.84 in Fig. 4) was similar to that of the global DROZY model (0.85 in Fig. 5) despite the personalized model being transferred from the global YawDD model. These findings further support the effectiveness of personalization in the FedSWP-TL framework.

Future Work

(1) While our proposed method showed promising results, it only utilized a single visual data modality. Future research could explore the integration of multiple modalities, such as audio and physiological data, to further improve the accuracy of fatigue detection. Additionally, the methods of modal fusion could be investigated to determine the most effective way to combine the information from different modalities. This could involve using DL architectures designed explicitly for multi-modal data or developing novel fusion techniques.

(2) Due to limited computing servers for the experiment, FedSWP-TL learning was only trained and tested on a small number of SWP nodes. The researchers noted that further experimentation on a larger scale is necessary to determine the full potential of FedSWP-TL.

(3) Fatigue is a complicated physiological state; however, this study did not categorize mental fatigue into distinct levels such as mild, moderate, and severe. The application of triplet loss with positive and negative samples poses a challenge for multi-level fatigue recognition, as it necessitates a nuanced redefinition of distances between varying levels of fatigue. As the FedSWP-TL framework emphasizes high-order fatigue embedding representations, future research could investigate the utilization of decoders for downstream tasks that accommodate different fatigue manifestations and judgment criteria.

(4) Given annotating constraints such as limited number, time, and the efficiency of professionals, obtaining precise and objective annotations is usually unfeasible. Semi-supervised learning methods leveraging abundant under-labeled samples would be further elaborated for fatigue monitoring.

Conclusion

This paper introduces a lightweight and adaptive federated smart work package with triplet loss (FedSWP-TL) framework to address the need for efficient fatigue monitoring of crane operators. The FedSWP-TL approach was conducted on YawDD, simulated DROZY, and a real-construction dataset ConPPMF. The experimental results reveal that the superiority of FedSWP-TL in adaptivity to more diverse groups of data-insufficient SWPs (with new subjects using diverse sensors). They also demonstrate improved efficiency in monitoring metal fatigue status with fewer overhead while maintaining high accuracy, showcasing its potential for resource-constrained implementation. The study contributes innovative techniques for fatigue monitoring, including triplet loss-based representations, selective gradient compression, and a lightweight model design, enhancing efficiency and adaptability in monitoring mental fatigue. Given data-insufficient and resource-constrained SWPs at construction sites, it also contributes to a more flexible and interpretable mental fatigue monitoring system using classification prediction and similarity analysis. Future research directions include exploring the integration of multiple modalities, conducting experiments on a larger scale, and differentiating varying mental fatigue levels in downstream tasks and leveraging abundant under-labeled samples for fatigue monitoring.

Data Availability Statement

Both the YawDD and DROZY datasets are readily available online (Abtahi et al. 2014, Massoz et al. 2016), while the models and scripts used for data processing and model training during this study can be accessed upon reasonable request at https://github.com/CI3LAB.

Acknowledgments

The authors would like to appreciate some data provided by Shaohui Huang and Ming Li for further validation on real-time construction worker monitoring. The work described in this paper is supported by grants from the Research Grants Council of the Hong Kong SAR of China, the National Natural Science Foundation of China, The University of Hong Kong (RGC Project No. 15219422 & G-HKU502/22, NSFC Project No. 72201228, HKU Project No. 2201100548 & 109000053), and Guangdong-Hong Kong Technology Cooperation Funding Scheme (TCFS) (Ref No.GHP/321/22SZ).

References

Abtahi, S., Omidyeganeh, M., Shirmohammadi, S., & Hariri, B. (2014, March). "YawDD: A yawning detection dataset." *In Proc., 5th ACM Multimedia Syst. Conf.* (pp. 24-28). https://doi.org/10.1145/2557642.2563678

Al Imran, M. A., Nasirzadeh, F., & Karmakar, C. (2024). "Designing a practical fatigue detection system: A review on recent developments and challenges." *J. Safety Research, 90,* 100-114. https://doi.org/10.1016/j.jsr.2024.05.015

Cai, J., Gao, Z., Guo, Y., Wibranek, B., & Li, S. (2024). FedHIP: Federated learning for privacy-preserving human intention prediction in human-robot collaborative assembly tasks. *Adv. Eng. Inf., 60*, 102411. https://doi.org/10.1016/j.aei.2024.102411

Chou, J. S., & Liu, C. Y. (2025). "Optimized Lightweight Edge Computing Platform for UAV-Assisted Detection of Concrete Deterioration beneath Bridge Decks." *J. Comput. Civ. Eng.,* 39(1), 04024045. https://doi.org/10.1061/JCCEE5.CPENG-5905

Dosovitskiy, A. (2020). "An image is worth 16x16 words: Transformers for image recognition at scale." *arXiv Preprint*. https://doi.org/10.48550/arXiv.2010.11929

Guarda, L., Tapia, J. E., Droguett, E. L., & Ramos, M. (2022). "A novel Capsule Neural Network based model for drowsiness detection using electroencephalography signals." *Expert Systems with Appl., 201,* 116977. https://doi.org/10.1016/j.eswa.2022.116977

Hart, S. G. (2006, October). "NASA-task load index (NASA-TLX); 20 years later." *In Proc., Human Factors and Ergonomics Society Anal. Meeting* (Vol. 50, No. 9, pp. 904-908). Sage CA: Los Angeles, CA: Sage Publications. https://doi.org/10.1177/154193120605000909

599 Kaida, K., Takahashi, M., Åkerstedt, T., Nakata, A., Otsuka, Y., Haratani, T., & Fukasawa, K.
600 (2006). "Validation of the Karolinska sleepiness scale against performance and EEG
601 variables." *Clinic. Neurophys., 117(7),* 1574-1581. https://doi.org/10.1016/j.clinph.2006.03.011
602 Li, J., Zhang, C., Zhao, Y., Qiu, W., Chen, Q., & Zhang, X. (2022, June). "Federated
603 learning-based short-term building energy consumption prediction method for solving the
604 data silos problem." *In Building Simulation* (Vol. 15, No. 6, pp. 1145-1159). Beijing:
605 Tsinghua University Press. https://doi.org/10.1007/s12273-021-0871-y
606 Li, X., Chi, H. L., Lu, W., Xue, F., Zeng, J., & Li, C. Z. (2021). "Federated transfer learning
607 enabled smart work packaging to preserve the personal image information of construction
608 workers." *Autom. Constr., 128,* 103738. https://doi.org/10.1016/j.autcon.2021.103738
609 Li, X., Chi, H. L., Wu, P., & Shen, G. Q. (2020). Smart work packaging-enabled constraint-
610 free path re-planning for tower crane in prefabricated products assembly process. *Adv. Eng.*
611 *Inf., 43*, 101008. https://doi.org/10.1016/j.aei.2019.101008
612 Li, X., Chi, H. L., Zhang, W. F., & Shen, Q. G. (2019). "Monitoring and alerting of crane
613 operator fatigue using hybrid deep neural networks in the prefabricated products assembly
614 process." *In ISARC. Proc., Int. Symp. on Autom. and Robotics in Constr.* (Vol. 36, pp. 680-
615 687). IAARC Publications. https://doi.org/10.22260/ISARC2019/0091
616 Li, X., Yang, X., Ma, Z., & Xue, J. H. (2023). "Deep metric learning for few-shot image
617 classification: A review of recent developments." *Pattern Recognition, 138*, 109381.
618 https://doi.org/10.1016/j.patcog.2023.109381
619 Li, X., Zeng, J., Chen, C., Li, T., & Ma, J. (2024). Decentralized adaptive work package
620 learning for personalized and privacy-preserving occupational health and safety monitoring
621 in construction. *Autom. Constr., 165,* 105556. https://doi.org/10.1016/j.autcon.2024.105556
622 Lu, Y., Liu, C., Chang, F., Liu, H., & Huan, H. (2023). "JHPFA-Net: Joint head pose and
623 facial action network for driver yawning detection across arbitrary poses in videos." *IEEE*
624 *Trans. on Intell. Transp. Syst., 24(11),* 11850-11863. https://doi.org/10.1109/TITS.2023.3285923
625 Ma, Y., Zhao, S., Wang, W., Li, Y., & King, I. (2022). "Multimodality in meta-learning: A
626 comprehensive survey." *Knowledge-Based Syst., 250*, 108976.
627 https://doi.org/10.1016/j.knosys.2022.108976

628 Massoz, Q., Langohr, T., François, C., & Verly, J. G. (2016, March). "The ULg multimodality
629 drowsiness database (called DROZY) and examples of use." *In 2016 IEEE Win. Conf. Appl.*
630 *of Computer Vision (WACV)* (pp. 1-7). IEEE. https://doi.org/10.1109/WACV.2016.7477715
631 Mehmood, I., Li, H., Umer, W., Arsalan, A., Anwer, S., Mirza, M. A., ... & Antwi-Afari, M.
632 F. (2023). "Multi-modal integration for data-driven classification of mental fatigue during
633 construction equipment operations: Incorporating electroencephalography, electrodermal
634 activity, and video signals." *Developments in the Built Environment, 15*, 100198.
635 https://doi.org/10.1016/j.dibe.2023.100198
636 Mehmood, I., Li, H., Umer, W., Arsalan, A., Shakeel, M. S., & Anwer, S. (2022). "Validity of
637 facial features' geometric measurements for real-time assessment of mental fatigue in
638 construction equipment operators." *Adv. Eng. Inf., 54,* 101777.
639 https://doi.org/10.1016/j.aei.2022.101777
640 Mehta, S., & Rastegari, M. (2021). "Mobilevit: lightweight, general-purpose, and mobile-
641 friendly vision transformer." *arXiv Preprint*. https://doi.org/10.48550/arXiv.2110.02178
642 Michielsen, H. J., De Vries, J., & Van Heck, G. L. (2003). "Psychometric qualities of a brief
643 self-rated fatigue measure: The Fatigue Assessment Scale." *J. Psychosomatic Research,*
644 *54(4)*, 345-352. https://doi.org/10.1016/S0022-3999(02)00392-6
645 Moriai, S. (2019, June). "Privacy-preserving deep learning via additively homomorphic
646 encryption." *In 2019 IEEE 26th Symp. on Computer Arithmetic (ARITH)* (pp. 198-198).
647 IEEE. https://doi.org/10.1109/ARITH.2019.00047
648 Subramanyan, P., Sinha, R., Lebedev, I., Devadas, S., & Seshia, S. A. (2017, October). "A
649 formal foundation for secure remote execution of enclaves." *In Proc., 2017 ACM SIGSAC*
650 *Conf. Computer &Comm. Security* (pp. 2435-2450). https://doi.org/10.1145/3133956.3134098
651 Sun, Z., Miao, Y., Jeon, J. Y., Kong, Y., & Park, G. (2023). Facial feature fusion
652 convolutional neural network for driver fatigue detection. Eng. App. of Artificial Intell.,
653 126, 106981. https://doi.org/10.1016/j.engappai.2023.106981
654 Wang, L., Li, H., Yao, Y., Han, D., Yu, C., Lyu, W., & Wu, H. (2023). "Smart cushion-based
655 non-invasive mental fatigue assessment of construction equipment operators: A feasible
656 study." *Adv. Eng. Inf., 58,* 102134. https://doi.org/10.1016/j.aei.2023.102134

Wu, H. T., Li, H., Chi, H. L., Kou, W. B., Wu, Y. C., & Wang, S. (2024). A hierarchical federated learning framework for collaborative quality defect inspection in construction. *Eng. App. of Artificial Intell.*, 133, 108218. https://doi.org/10.1016/j.engappai.2024.108218

Guo, Y., Wang, C., Yu, S. X., McKenna, F., & Law, K. H. (2022). Adaln: a vision transformer for multidomain learning and predisaster building information extraction from images. *J. Comput. Civ. Eng., 36(5)*, 04022024. https://doi.org/10.1061/(ASCE)CP.1943-5487.0001034

Xu, Y., Jiang, Z., Xu, H., Wang, Z., Qian, C., & Qiao, C. (2023). "Federated learning with client selection and gradient compression in heterogeneous edge systems." *IEEE Trans. on Mobile Computing,23(5)*,5446 - 5461. https://doi.org/10.1109/TMC.2023.3309497

Yan, Y., Mao, Y., Shen, Z., Wei, Y., Pan, G., & Zhu, J. (2021). "A high-efficiency fatigued speech feature selection method for air traffic controllers based on improved compressed sensing." *J. Healthcare Eng., 2021*, 1-10. https://doi.org/10.1155/2021/2292710

Yang, H., Ge, M., Xue, D., Xiang, K., Li, H., & Lu, R. (2023). "Gradient leakage attacks in federated learning: Research frontiers, taxonomy and future directions." *IEEE Network,38(2)*, 247 - 254. https://doi.org/10.1109/MNET.001.2300140

Yang, L., Yang, H., Wei, H., Hu, Z., & Lv, C. (2024). "Video-based driver drowsiness detection with optimised utilization of key facial features." *IEEE Trans. on Intell. Transp. Syst.,25(7)*, 6938 - 6950. https://doi.org/10.1109/TITS.2023.3346054

Zhu, Q., Wang, H., Xu, B., Zhang, Z., Shao, W., & Zhang, D. (2022). "Multi-modal triplet attention network for brain disease diagnosis." *IEEE Trans. on Medical Imaging, 41(12),* 3884-3894. https://doi.org/10.1109/TMI.2022.3199032

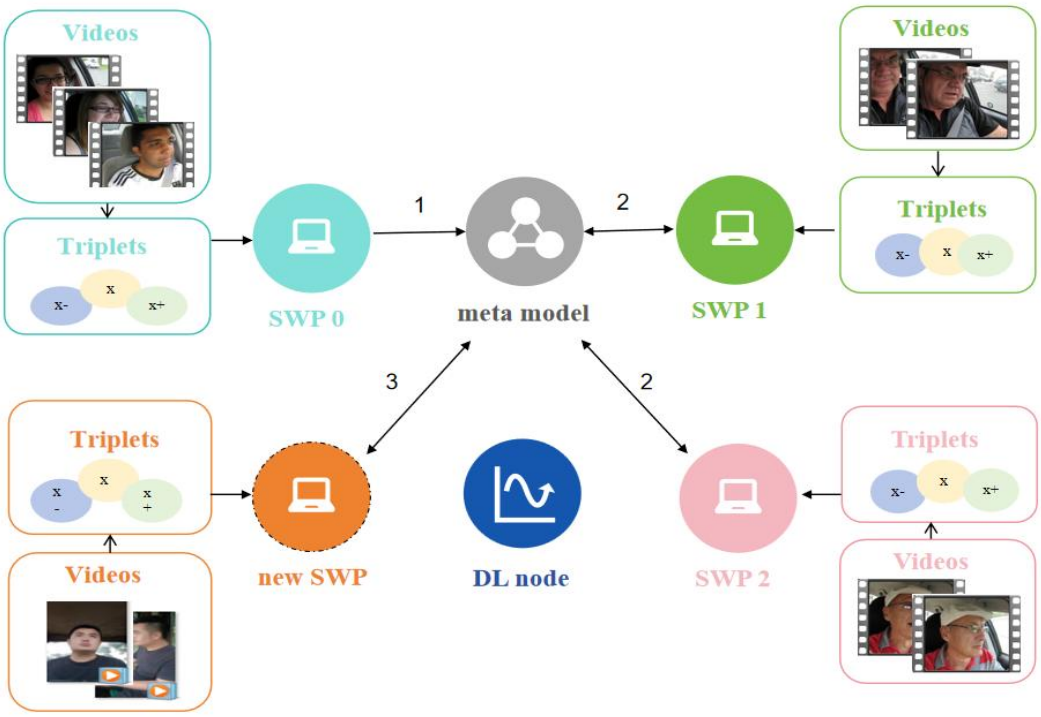

Fig.1. The FedSWP-TL network

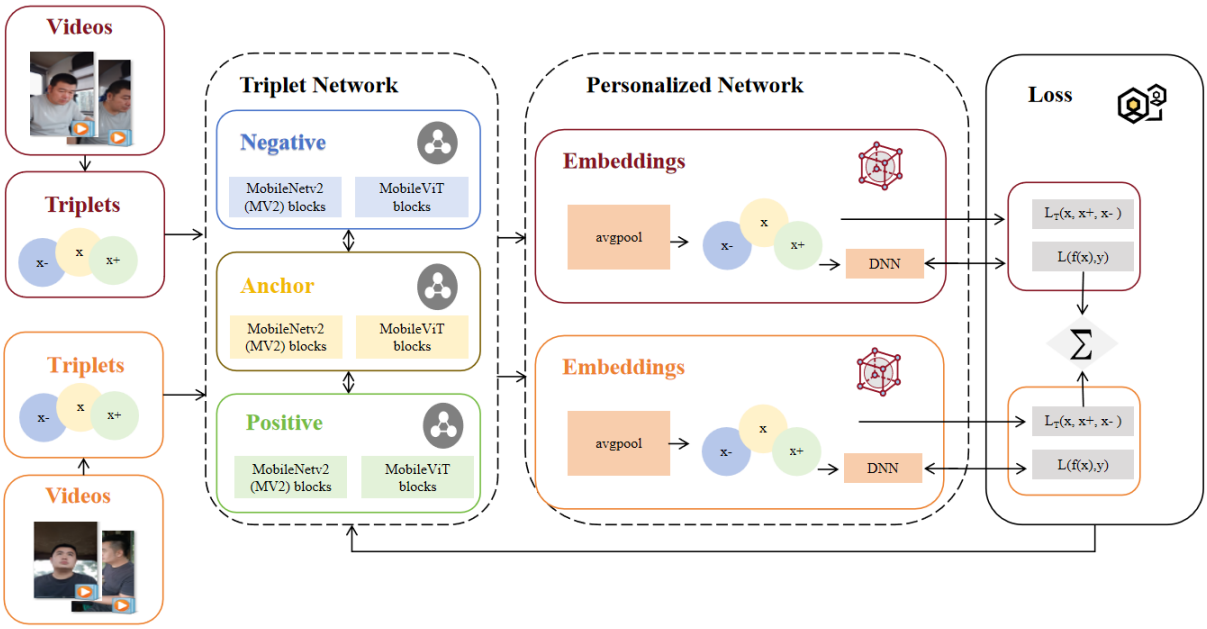

Fig.2. Training process of FedSWP-TL

Fig.3. Sample frames of SWP monitoring videos

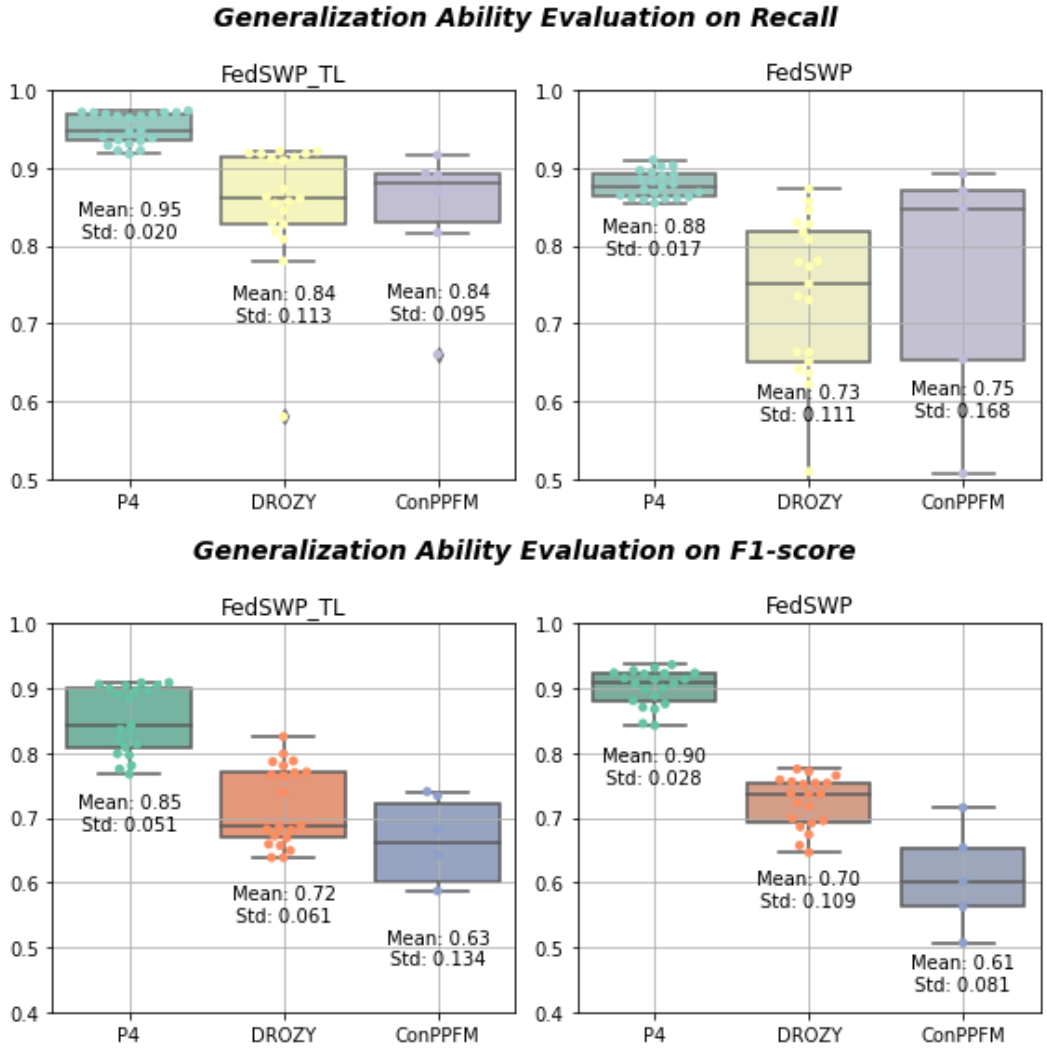

Fig.4. Generalization ability evaluation

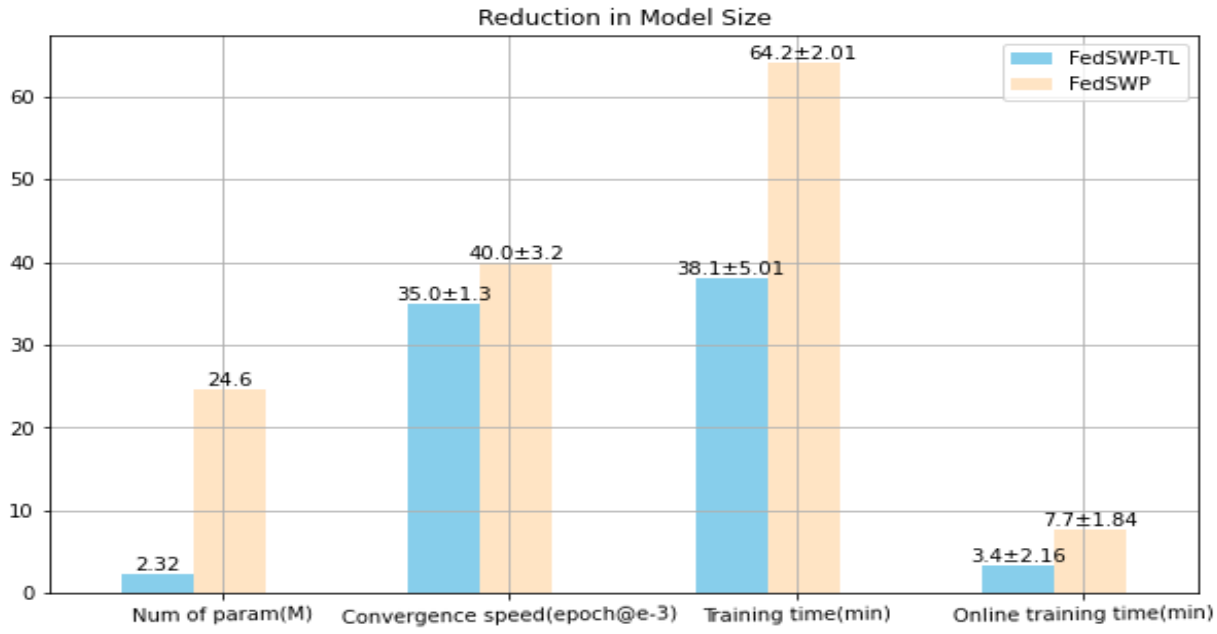

Fig.5. Resource overhead

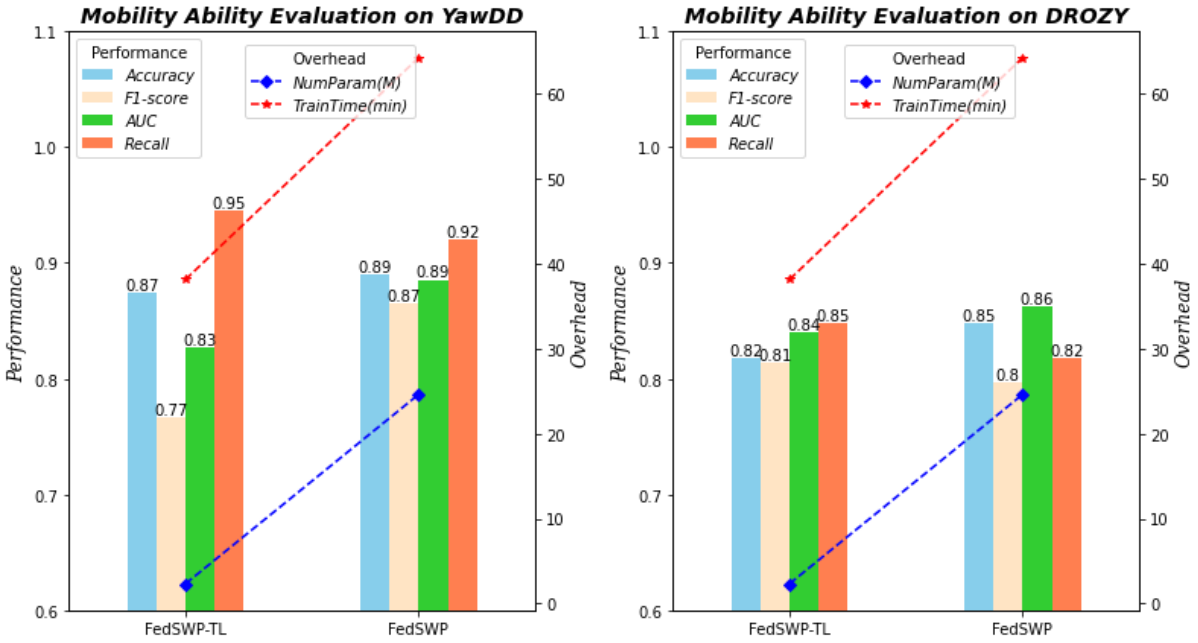

Fig.6. Performance vs resource overhead in SWP0

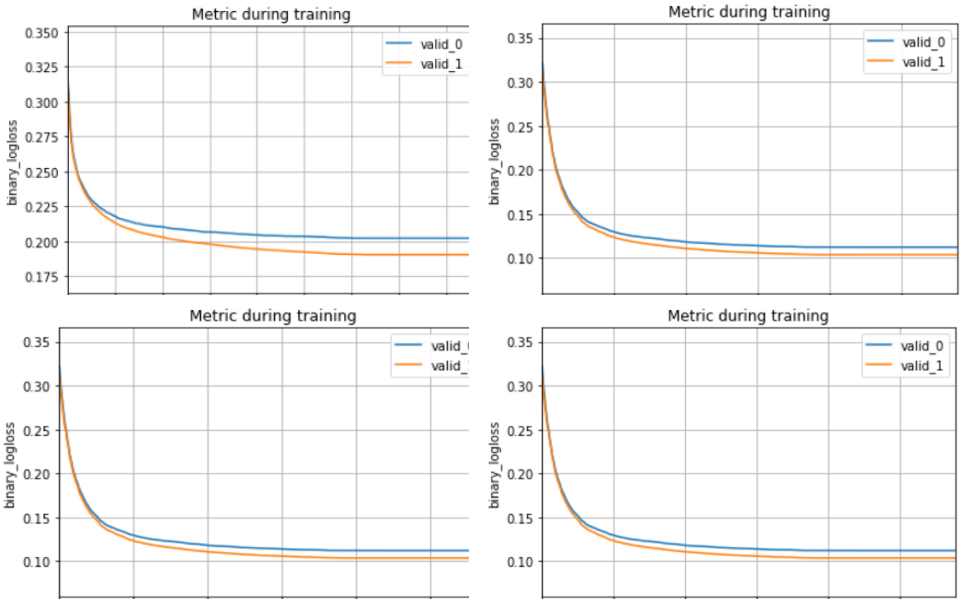

Fig.7. Loss curves in SWP0, SWP1, SWP2, and SWP3

Table 1. The architecture of the SWP model in FedSWP-TL

Input	Operator	Share	C, n, s
256×256×3	conv2d	shareable	16,1,2
128×128×16	MV2	shareable	32,1,2
128×128×32	MV2, ↓2	shareable	48,1,4
64×64×48	MV2	shareable	48,2,4
64×64×48	MV2, ↓2	shareable	64,1,8
32×32×64	MobileViT block	shareable	64,1,8
32×32×64	MV2, ↓2	shareable	80,1,16
16x16x80	MobileViT block	shareable	80,1,16
16x16x96	MV2, ↓2	shareable	96,1,32
8×8×96	MobileViT block	shareable	96,1,32
8×8×96	conv2d 1x1	shareable	384,1,32
8×8×384	avgpool 8x8	shareable	384
384	Dense	personalized	64
64	Dense	personalized	2

Table 2. Summary of datasets

Datasets	Subjects	Age	Gender	Colors	Sensors
YawDD	90	20-30	47 males, 43 females	Caucasian, African, Middle-eastern, Asian	RGB cameras for faces from the front mirror
DROZY	14	20-25	3 males, 11 females	European	NIR cameras and an Embla Titanium system
ConPPMF	3	40-50	3 males	Asian	RGB cameras for faces and smartwatches

Table 3. Processed data statistics on YawDD

SWP Nodes	Total workers	Total Frames	Frames train	Frames test	Fatigue % test	Triplets train	Fatigue % train
SWP 0	66	49,182	6,067	1,617	26.5%	14,776	37.2%
SWP 1	5	3,124	410	102	33.4%	960	34.1%
SWP 2	5	4,350	573	143	30.1%	1,304	27.6%
SWP 3	5	3,513	462	115	31.3%	1,045	26.2%

Table 4. The hyper-parameter in FedSWP-TL

Hyper-parameter	Explanation/Usage	Scope	Optimal values
Learning rate	Governing the step size taken at each iteration while the model moves closer to the minimum of its loss function.	{1e-5,1e-3,5e-3,1e-2,5e-2}	0.01
Batch size	Determining the number of samples processed in a single iteration during training	$\{2^k$ k=1,2…12}	512
Epoch	Representing the number of times the model processes the entire training dataset during training	{5, 10, 20, 25, 50}	25
Dropout	Reducing the interdependence of neurons and enhancing the model's ability to generalize and prevent overfitting.	{10%, 20%, 30%…60%}	50%
Activation function	Introducing non-linear transformations within the model, enabling it to capture complex data patterns and relationships	{sigmoid, Relu, elu}	Relu
Optimizer	Adjusting the model's parameters during training to minimize the loss function	{Adam, RMSprop, Nadam, sgd, adagrad}	Adam
Number of SWPs	Number of SWPs in the training network	{2,4,6}	4

Table 5. Data statistics of new SWP Nodes

SWP Nodes	Total workers	Original Resolution	Total Frames	Frames test	Fatigue % test
P4	8	(480 × 640 × 3)	5,190	5,203	23.9%
DROZY	14	(424 × 512× 3)	66,879	10,025	12.7%
ConPPMF	2	(1280 × 720 × 3)	122,663	886	21.8%

Table 6. SWPs' model results on YawDD

	Accuracy	F1-score	AUC	Recall
SWP1	0.895 ±0.004	0.785 ±0.00	0.901 ±0.00	0.855 ±0.00
SWP2	0.904 ±0.003	0.814 ±0.00	0.907 ±0.00	0.894 ±0.00
SWP3	0.913 ±0.005	0.786 ±0.00	0.896 ±0.00	0.863 ±0.00
SWP0'	0.817 ±0.001	0.767 ±0.00	0.892 ±0.00	0.878 ±0.00

Algorithm 1. The Process of the Proposed FedSWP-TL.

Algorithm 1: FedSWP-TL

Input: SWP triplet sets $\{T_k = \mid k \in K\}$ from historical datasets $D = \{\mathcal{D}_k\}_k^K$;
Output: global parameters Θ_{meta}; And personalized weight $\theta_1, \theta_2 \ldots \theta_K$ for SWP 1-K

Cloud Execution:
Initialize MobileVit parameters $\Theta_{meta} \leftarrow \min \mathcal{L}(T_0|\Theta_{meta})$, T_0 from a public dataset D_0
For each round, t=1,2,3, ... do
    Select $C(C \leq K)$ SWPs that finished local update at t-1
    Determine the compression ratio r^k for each SWP
    Each SWP executes $SWPupdate(\Theta_{t-1}, r^k)$
    $\Theta_t \leftarrow 1/C \sum_k^C \alpha^k SWPupdate(\Theta_{t-1}, r^k)$
Return Θ_t
    Each $SWPupdate(\Theta_{t-1}, r^k)$:
    $\theta_{t-1} \leftarrow \{\Theta_{t-1}, \theta_k\}$
    $\delta \leftarrow f(r^k)$
    For local updates with triplets T_k:
        $\mathcal{L} \leftarrow \mathcal{L}_T(\boldsymbol{x}, \boldsymbol{x}_+, \boldsymbol{x}_-) + \mathcal{L}(f(\boldsymbol{x}), \boldsymbol{y})$
        $\nabla\theta_t \leftarrow \nabla\mathcal{L}$
        $|\nabla\theta_t|^\delta \leftarrow$select gradient element with absolute value $>\delta$ after error compensation
        accumulate error for gradient elements that are not selected
    Return $|\nabla\theta_t|^\delta$