A privacy-preserving multi-task learning framework for emotion and identity recognition from multimodal physiological signals

Mohamed Benouis, Yekta Said Can, Elisabeth André

Angaben zur Veröffentlichung / Publication details:

Benouis, Mohamed, Yekta Said Can, and Elisabeth André. 2023. "A privacy-preserving multi-task learning framework for emotion and identity recognition from multimodal physiological signals." In *2023 11th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW), 10-13 September 2023, Cambridge, MA, USA*, 1–8. Piscataway, NJ: IEEE. https://doi.org/10.1109/aciiw59127.2023.10388160.

A Privacy-Preserving Multi-Task Learning Framework For Emotion and Identity Recognition from Multimodal Physiological Signals

1st Mohamed Benouis
Chair for Human-Centered AI
Universität Augsburg

2nd Yekta Said Can
Chair for Human-Centered AI
Universität Augsburg

3rd Elisabeth André
Chair for Human-Centered AI
Universität Augsburg

Abstract—**The increasing popularity of empathetic sensors can play a significant role in the affective computing era. Recognizing human emotion from these unobtrusive devices is an important building block in this context. Multi-task learning has been studied extensively for various machine learning tasks, including affective computing, which uses the shared information between related tasks to improve performance. Since the physiological data from the mentioned sensors contain private data, they can also lead to privacy threats by exposing highly sensitive information. To address this issue, we combine differential privacy and federated learning approaches with multi-task learning to efficiently recognize the user's mental stress while perturbing private user identity information. More concretely, the proposed framework improves the performance of emotion recognition tasks by taking advantage of multi-task learning and preserving privacy. We extensively evaluate our framework with the employed dataset: results show an accurate emotion recognition of 90% while limiting the re-identification accuracies to 47%.**

I. Introduction

Affect recognition is a significant research field that has a direct impact on our daily lives. Affect recognition research makes use of facial expressions [1] and speech [2] because the correlation between facial expressions and speech with emotions is known for a long time. The physiology-based solution is another alternative for affect recognition. Its main advantage is its suitability for continuous monitoring in daily life without disturbing users. Wearable devices are pervasive tools for collecting passive and quantitative physiological data. More than 330 million smartwatches, fitness trackers, and similar wearables have been sold, and the market has been growing each year [3]. Physiological, environmental, and activity-related data can be collected without interrupting users, which makes them a favorable prospect for recognizing affects.

Multi-task learning (MTL) is proposed to address the training of multiple related tasks simultaneously. MTL transfers knowledge between these multiple tasks to improve the performance of each model [4]. It can be considered as an implicit data augmentation technique or eavesdropping of additional supervision to enhance the generalization capability of machine learning (ML) models [5]. MTL is applied to improve the performance of ML models in various domains including affective computing [6].

Nevertheless, collecting and analyzing personal private data, such as identity, gender, age, and user location, raise concerns about users' privacy, especially when the analysis is performed in a cloud-based wearable application. For instance, cloud-based ML algorithms can provide beneficial services (e.g., stress monitoring or health apps), but the redundant data collected from the users could be used for illegitimate purposes (e.g., identity recognition for targeted social advertising). Intuitively, extracting such information without an individual's consent can be considered a violation of their privacy. In this regard, researchers often employ data obfuscation techniques to anonymize private data users; however, most of them rely only upon relational data and cannot be applied to model training data. In addition, they often require human intervention to label the sensitive information held in the original data, which is usually computationally expensive.

Recent research has explored the possibility of modifying the learning paradigm such that feature attributes may not be shared while the only learning model can be used only for legitimate tasks. Researchers proposed a Federated Learning (FL) approach to train the user data locally. It allows the users to train their data collaboratively by sharing trained model parameters instead of original data. It abides by the data protection laws such as EU General Data Protection Regulation (GDPR) and California Consumer Privacy Act (CCPA). The emerging concept of FL has shown promising solutions for preserving user privacy in healthcare networks, social networks, and the IoT.

The FL approach is a big step in preserving users' privacy but is not invincible. Sensitivity information can still be retrieved by applying reverse engineering to the local model parameters. On top of FL, researchers applied privacy-preserving approaches in ML for physiological data. Differential privacy (DP) is a well-known method that has been largely used in an FL paradigm to mitigate such a trade-off. It injects noise into each model in the client or server, perturbs the updates, and limits gradient leakage shared between the nodes (i.e., client and server) [7].

In practice, protecting the user's privacy without degradation in model utility is still an open problem. In this study, we first

Corresponding author: Mohamed Benouis, mohamed.benouis@uni-a.de

This work was carried out within the framework of the AI Production Network Augsburg.

TABLE I: Studies using MTL or Privacy preserving approaches for various applications.

Study	Application	Signal	MTL	FL	DP	Multimodal
[10]	Network Anomaly Detection	Network Signals	✓	✓	X	X
[11]	Activity recognition	Acceleration, Gyroscope	✓	✓	X	✓
[12]	Affect recognition	Speech, video, text	X	✓	X	✓
[9]	Affect recognition	Face images	✓	X	X	X
[13]	Affect recognition	Face images	X	✓	X	X
[14]	Affect recognition	Face and speech	X	✓	X	✓
[15]	Affect recognition	Speech	X	✓	✓	X
[16]	Affect Recognition	PPG	X	✓	X	X
Our Study	Affect Recognition	PPG, EDA, Acceleration, ST	✓	✓	✓	✓

implemented a multitask learning architecture for recognizing stress and identity from multimodal physiological signals. We further added FL and DP mechanisms to preserve privacy. In this way, we were able to improve the stress recognition performance with the help of MTL and hide identity information by adding noise to the identity task model by using DP. To the best of our knowledge, this study is the first MTL-based affect recognition study using FL and DP to preserve privacy at the same time.

II. Related works

In order to develop a robust affective computing system that can be used in practical applications, researchers tested various modalities with the state of the art deep learning techniques. Multi-task learning has also raised significant attention from various domains over the past few years, including affective computing. It was applied to audiovisual signals [8] for detecting arousal and valence levels in a continuous manner. In another study, MTL was used for detecting smile detection, emotion recognition and gender classification [9]. By using MTL with CNN, they achieved better accuracies on benchmark datasets. However, when private tasks (i.e. face, gender, person detection) are included in MTL to improve affect recognition performances, the models create the risk to reveal this sensitive information to possibly malicious parties.

Data privacy has become an issue of great concern in affect recognition using either verbal or nonverbal data, as the gender, age, and identity of the user could be revealed in the process. For instance, statistical manipulation could exploit facial images of users to assume their identities and infiltrate biometric authentication systems. FL is proposed to preserve privacy while taking advantage of ML. It attracted significant attention from various domains over the past few years, affective computing research and applications on emotion recognition-related tasks are rarely discussed. Most existing works are conducted on private datasets or in limited scenarios, making it difficult for researchers to compare their methods and push the frontier forward fairly. Arji et al. [17] used an FL approach to train N (i.e., number of clients) local feed-forward neural network model using multimodal streaming data, called DEAP dataset to predict the underlying valence-arousal level and monitor the emotion status of the users in real-time. Their approach has been validated on their own datasets and achieved a better average accuracy of 0.842 %. In [18], they investigated FL with the two affective computing tasks of classifying self-report and perception ratings. This approach was developed to classify affective constructs from self-reports and perception ratings in audio, video, and text datasets. In another context related to the facial expression task, in [13], they trained a few-shot federated self-learning framework on facial expression with partially annotated data. Their approach has been validated on two datasets and achieved an accuracy of 84.9 % on the first and 97.3% accuracy on the second dataset, respectively. Using speech modality, in [19], they investigated an FL approach in emotion recognition tasks while sharing only the model among clients. Their approach has been validated on the EMOCAP dataset and achieved 54.8 unweighted average recall (UAR) (%) using the LSTM classifier. In another work [14], they combined face and speech modalities using the FL approach. Their proposed framework has been validated and tested for facial and speech emotion recognition on their own dataset, and it achieved an accuracy of 71.64% and 85.04%, respectively. On the other hand, FL is used rarely for recognizing affects from physiological signals. Can and Ersoy [16] employed the FL learning model to predict stress mood using physiological data. Each sub-client performs an MLP classifier to train its own local data on edge and the sharing of each single updating parameters MLP using the FedAVG algorithm.

Although FL has been proposed to improve the training model in term of privacy, the privacy vulnerabilities of the stochastic gradient descent (SGD) algorithm remains unsolved. Moreover, since affective recognition task often handles important labeling and annotation data that attackers may use to infer sensitive information, such as age, gender, identity, and so on. DP mechanism has been mostly spoken in FL settings, and it implies injecting noise into each model client or server, perturbing the updates, and limiting gradient leakage shared among nodes (i.e., client and server) [20]. In one of the first applications, authors introduced a new private training method, called differential private stochastic gradient descent, that decreases local and global gradient information leakage between the client and server. Instead of using the standard composition theorem to compute the final distribution of overall noise clients, they used a moment's accountant metric to adaptively track the overall privacy loss. Since the servers are often curious or untrustworthy, Wei et al. [21] introduced a local DP mechanism algorithm by adding distribution Gaussian noise into the user models before uploading them to servers. To scale the communication number required for optimal convergence upper bound for the DP, they introduced a new approach called a communication rounds discounting (CRD) method that can achieve a much better trade-off between the computational complexity of searching and the convergence performance. DP has been also used for affect recognition. Feng et al. [15] used user-level differential privacy (UDP) for alleviating privacy leaks of FL for speech emotion recognition. Recently, Smith et al. [11] demonstrated a promising performance for preserving privacy via multi-task FL for activity recognition.

When we examine the affective computing literature, although there are studies using MTL to improve performance and using FL or DP to preserve privacy separately, there is no study that both takes advantage of MTL and preserves privacy. This study is the first MTL-based affect recognition study using FL and DP to preserve privacy at the same time.

III. Proposed Methodology

Our proposed framework is divided into three main substeps: feature extraction, FL model, and FL with DP settings.

A. Data Description and Feature Extraction

WESAD dataset [22] has been created for affective state monitoring. Each participant recorded physiological signals such as blood volume pulse, electrocardiogram, electrodermal activity, electromyogram, respiration, body temperature, and three-axis acceleration measured from the chest and wrist using RespiBAN and Empatica E4 devices. Fifteen people (12 males and three females) participated in this experience. There were four states: baseline, amusement, stress, and meditation. More details can be found in [22]. Each modality signal is segmented using 700 sample windows size with 50% overlap, as suggested in the literature [23]. In total, 121813 segments were created. To maximize the correlation among inter-subjects and minimize among subjects, these segments were further processed for extracting features [24] such as mean, variance, root mean square, frequency domain features, average first amplitude difference, second amplitude difference, skewness, kurtosis, and entropy as a nonlinear feature.

B. Decentralized multi-task FL model

For privatizing the user's identity while preserving stress recognition accuracy, we adopted a multi-task FL approach that can effectively improve the performance of stress recognition while limiting the risk of inferring sensitive information from the training model since the client does not want to be exposed to the cloud service provider. Multi-task FL architecture-based stress recognition is developed as follows:

1) The dataset is partitioned into k clients. The data size of all the clients is the same. The client distribution is also assumed as independent and identically distributed (IID) and not independent and identically distributed (No-IID) [25].
2) For the local training process, only one iteration for SGD local training for each client. In particular, w is the local model parameter [25], given by:

$$\mathbf{w}_U^{D_i} = \arg\min_{\mathbf{w}_U} \quad F_U\left(\mathbf{w}_U\right) + \frac{\mu}{2} \quad \mathbf{w}_U - \mathbf{w}^{(D_i-1)} ||^2 \Bigg) \tag{1}$$

3) The local data of different clients cannot be communicated and only the models can be shared.
4) The server employs a global averaging approach to aggregate all local training models to compute the final global model. Formally [25], the server aggregates the weights sent from the K clients as (FedAvg), as:

$$\mathbf{w} = \sum_{U=1}^{K} p_i \mathbf{w}_U^{D_i} \tag{2}$$

Where w_i is the parameter vector trained at the k^{th} client, w is the parameter vector after aggregating at the server, K is the number of participating clients, D_i is the dataset size of each participating client, $D = \bigcup D_i$ the whole distributed dataset, and $P_U = |\mathcal{D}_i| / |\mathcal{D}|$.
5) The global training epoch is set to M rounds (aggregations). The server solves the optimization problem [25]

$$\mathbf{W}^* = \arg\min_{\mathbf{w_U}} \sum_{U=1}^{M} P_U F_U\left(\mathbf{w_U}, \mathcal{D}_i\right) \tag{3}$$

where F_U is the local loss function of the k^{th} client. Generally, the local loss function is given by local empirical risks.

C. Decentralized FL with DP

In conventional FL, the global model is computed through averaging over model client participants, which performs better within homogeneous FL settings. However, employing inference or adversarial attack, this shared model may contain sensitive and private information such as gender, age, biometric template user, etc. In such cases, the MFL framework is required to reduce the leakage of the black box gradient exchanged model. To overcome this limitation, researchers have employed the DP scheme to protect either local or global data training FL model. However, the perturbed gradient using DP with a low budget has high variance, which leads to worse performance and slower convergence. Motivated by personalized FL [26], our work focuses on client-level privacy, which aims to a private specific layer of the client model rather than perturbing the entire whole local model. This is because the base layers are mostly redundant information, while the most important information that holds private and public information is located in the upper layer. To meet the utility privacy trade-off guarantee for the personalization FL model, the DP mechanism is to perturb the gradients using Gaussian noise at a specific layer or task. Here, We employ all steps in the FL model; except step 4, i.e., before uploading the local SGD model client to the global server, we inject an amount of noise to the updated local parameters. In that sense, we will perturb the local gradient training inference with two kinds of noise distributions:

1) An additive Gaussian noise $\eta_{\sim}\mathcal{N}\left(0, \sigma_l^2\right)$ to each weight local model. This operation can be mathematically described as follows:

$$w_{t+1} = w_t + \eta \tag{4}$$

2) A set of noise distributions can be sampled from the DP mechanism (DP). A randomized mechanism M on the training set with domain $\mathcal{X}$ and range R satisfies

$(\epsilon, \delta)-\mathrm{DP}$ for two small positive numbers and if the following inequality holds [20]:

$$\operatorname{Pr}[\mathcal{M}(x) \in \mathcal{S}] \leq e^{\epsilon} \operatorname{Pr}\left[\mathcal{M}\left(x^{\prime}\right) \in \mathcal{S}\right]+\delta \tag{5}$$

where x and $x^{\prime} \in \mathcal{X}$ are two input neighbor datasets, and $S \subseteq R$ where R is the set of all possible outputs, δ is privacy loss or failure probability and ϵ is privacy budget.

An ideal DP mechanism provides a lower value of δ and a smaller value of ϵ. Unfortunately, these values decrease the function utility (e.g., accuracy metric), so the main question is how much DP values we must perturb its output while guaranteeing trade-off privacy-utility. Intuitively, an output perturbation mechanism takes an input x and returns a random variable $s(x)$. This operation can be modeled by:

$$\mathrm{M(x)}=\mathrm{s(x)}+N_{\sigma} \tag{6}$$

where N is scaling noise sampled from specific distribution. In this work, we chose Laplace and Gaussian mechanisms [20] that employ L1 and L2 norm sensitivity, respectively. The sensitivity function can be expressed as:

$$\Delta f=\max_{D, D^{\prime}} \quad s(x)-s\left(x^{\prime}\right)_{1,2} \Big\| \tag{7}$$

And scaling noise can be computed as:

$$\sigma=\Delta f / \varepsilon \tag{8}$$

Output perturbation satisfies $(\epsilon, \delta)-\mathrm{DP}$ when we properly select the value scaling noise. Thus, it sampled from Laplace and Gaussian distributions [20] as:

$$M_{\text{Laplace}}(x, f, \varepsilon, \delta)=s(x)+\operatorname{Lap}(\mu=0, b) \tag{9}$$

$$M_{\text{Gaussien}}(x, f, \varepsilon, \delta)=f(x)+\mathcal{N}\left(\mu=0, \sigma^{2}\right) \tag{10}$$

The gradient information leakage can be reduced by applying gradient thresholding or clipping algorithm. As explained in [7], gradient clipping is crucial in ensuring the DP of FL algorithms. So, each provider's/client's model update needs to have a bounded norm, which is ensured by applying an operation that shrinks individual model updates when their norm exceeds a given threshold. Clipping impacts of an FL algorithm's convergence performance should be known to create FL algorithms that protect DP.

IV. Experimental Results

Three scenarios are created to tackle the aforementioned challenges with the DP learning approaches: centralized, decentralized FL, and decentralized FL. Their performances are evaluated on the WESAD dataset, which consists of multi-modal physiological data of 15 individuals on two different tasks. The first task is identifying users from a set of registered and recorded users. The second task is perceived binary stress recognition, which tries to distinguish the user's stress level, stress vs. non-stress. We trained a multi-task deep learning model for simultaneous tasks. The accuracy metric is used for measuring identification and stress recognition performance. In each simulation scenario, we run 5-fold cross-validation, where each fold is tested based on the training of the other four. As described in Table 2, the multi-task 1D-CNN model is based on 3 convolutional layers, a pooling layer, 2 fully connected layers, and 2 linear classifiers to classify the studied tasks. The multi-task model uses the cross-entropy loss function and SDG Learning rate (β=0.0005).

Algorithm 1 Multi task FL approach with DP.

```
Number of communication rounds M, the initial global parameter $\boldsymbol{w}^0$, the sample ratio $q=\frac{K}{N}$, the clipping threshold C, the variance of noise $\sigma^2$ and DP parameter initializations $(\epsilon_i, \delta_i)$
$Initialize: t=0$
The server broadcasts w and T to all selected clients
while $t < M$ do
    for $i \in K$ do
        Update the local parameters as:
        $\mathbf{w}_i^{(t)}=\arg\min_{\mathbf{w}_i}\left(F_i(\mathbf{w}_i)+\frac{\mu}{2}\left\|\mathbf{w}_i-\mathbf{w}^{(t-1)}\right\|^2\right)$
        Clip the updated parameters model
        $\mathbf{w}_i^{(t)}=\mathbf{w}_i^{(t)}/\max\Bigg(1,\frac{\mathbf{w}_i^{(t)}}{C}$
                ▷ Perturb selected layers (full, shared, task)
     ▷ Add with DP or Gaussian noise, i.e., $\eta\sim\mathcal{N}\left(0,\sigma_l^2\right)$
        $\tilde{\mathbf{w}}_i^{(t)}=\mathbf{w}_i^{(t)}+\mathbf{n}_i^{(t)}$
    end for
    Update the global parameters $\boldsymbol{w}^{t+1}=\sum_{i\in\mathcal{K}}p_i\widetilde{\boldsymbol{w}}_i^{t+1}$
    The server broadcasts the global parameters

    for $\mathcal{C}_1,\mathcal{C}_2,\mathcal{C}_N$ do
        Test aggregating parameters on the local data
    end for
    $t=t+1$
end while
Results : $\widetilde{\mathbf{w}}^{(T)}$
```

TABLE II: Parameters of multi-task 1D-CNN model. Abbreviations are listed: IN: Input, OUT: Output, K: Kernel size(K).

Layer	Type	Parameters
Input	Input	Features size
Conv_1	Convolution	IN=1, OUT=20, K=8 Stride=1 , Padding
Relu_2	Activation function	Relu
Pooling_3	Pooling	S=2, Stride=2: Max Pooling
Conv_4	Convolution	IN=20, OUT=40, K=8 Stride=1 , Padding
Relu_5	Activation function	Relu
Pooling_6	Pooling	S=8, Stride=2: Max Pooling
Conv_7	Convolution	IN=40, OUT=60, K=8 Stride=1 , Padding
Relu_8	Activation function	Relu
Pooling_9	Pooling	S=2, Stride=2: Max Pooling
FC_1	Fully connected layer	IN=360, OUT=100
FC_2	Fully connected layer	IN=100,OUT=300
Linear_1	IN=100, OUT=2	Output=2 classes, activation function:linear
Linear_2	IN=300, OUT=15	Output=15 classes, activation function:linear

For each target task, the individual loss is determined by the cross-entropy for both stress recognition ($Loss_1$) and identification tasks ($Loss_2$). The individual losses are summed and form the total cost function ($Loss_T$).

A. Centralized learning (CL) approach

We carry out the CL approach on the WESAD data set as a baseline experiment. Here, only one training model

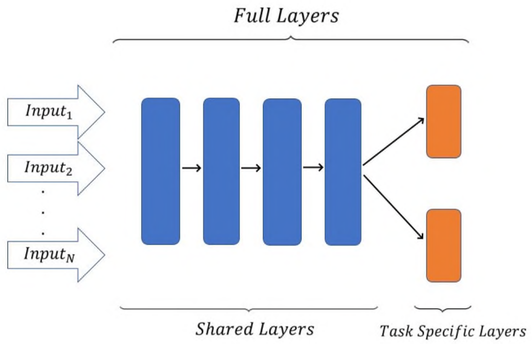

Fig. 1: Three different scenarios for adding noise with DP. Adding noise to only shared layers, task-specific layers and full layers.

was created to train and test the whole dataset. We set the maximum number of training epochs to 30.

Figure 2 shows the results of our centralized learning approach: a 1D-CNN multi-task for stress and identity recognition tasks. After training, the output layers are used to infer the stress mood and identity of the user, resulting in quite similar average scores of 99%. The results demonstrate the potential information leakage in this case, which preserves model accuracy while not protecting user privacy. As a result, this approach could not guarantee users' privacy since the user data is transmitted to the server, and the model is trained on the server side.

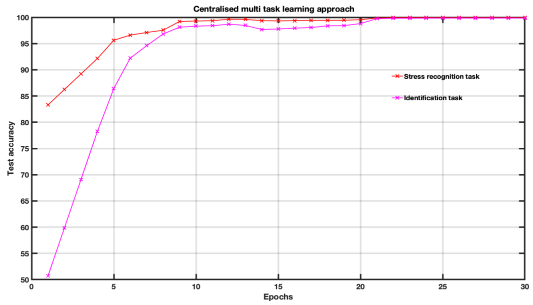

Fig. 2: Stress vs. identity recognition using multi-task FL

B. Multi task FL approach

Here, K (i.e., participating clients) training models were created to train the whole dataset and the size of local samples D_i=1000. We set the number of training epochs (communication round) to T=40 and local training epochs to 1. Overall, The best accuracy result achieved is 97% for stress mood recognition and 93% for user identification.

To examine the effect of client participation within the multi-task FL model, we tried different numbers of clients, i.e., K=5, k=10, and K=20. As reported in [27] and confirmed in Figure 3), an increasing number of clients and more client participation provide better performance for MFL training. The client distribution is different in assessing the MFL model in real-world conditions. We compare the convergence performance of the MFL model under IID and NO IID (see Figure 3). We can note that the data distribution dramatically affects the quality of the FL training and obviously affects MFL's convergence performance.

Adjusting FL hyper-parameter settings results can achieve a better performance than the centralized learning approach; however, it may lead to a lower privacy level. As a result, SGD training may still reveal sensitive information about the client while exchanging the ML model with the global server.

C. Multi-task FL with DP approach

To highlight the benefits of our proposed approach, we examine the impact of injecting noise into the local client training network according to these three scenarios: the full layers, shared layers, and task-specific layers (see Figure 1). The employed noise is sampled via the following mechanisms: 1) Without the DP technique, the noise scale is drawn from Gaussian distribution, i.e., $\eta \sim \mathcal{N}\left(0, \sigma^2\right)$. The noise-added parameters model can prevent the privacy breach with an appropriate choice of variance.

2) With DP technique-based Laplace and Gaussian mechanisms, the noise scale is drawn from the output perturbation mechanism. DP parameters are computed at each local training round to generate appropriate noise injected from specific distributions (i.e., Laplace and Gaussian). Besides appropriate (ϵ,δ)-DP initialization, there are a few hyper-parameters to be tuned, such as the number of clients N, the number of maximum communication rounds T and the number of chosen clients K.

a) Laplace distribution [7] is computed as:

$$\sigma = \frac{\Delta f}{} \tag{11}$$

a) Gaussian distribution, two distributions are given by [7] and [21] respectively.

$$\sigma_1 = \sqrt{\frac{2 \log \frac{1.25}{\delta}}{\varepsilon}} \tag{12}$$

$$\sigma_2 = \frac{\Delta f \sqrt{2qT \log \big(\frac{1}{\delta}}}{\varepsilon} \tag{13}$$

where $\Delta f = \frac{2C}{U_i}$; $q = \frac{K}{N}$; and C= clipping threshold. We set the clipping factor to 1 and δ to 0.00001. Figures 4 and 5 show the accuracy comparison of adding Gaussian distribution levels into local training according to the three scenarios. Compared to the baseline scenario, i.e., no private mechanism, the perturbing share layer scheme with σ=0.1 and σ=0.3 only provides better results for utility tasks; however, the identification task reached an accuracy of around 86%.

In this case, the amount of noise drawn from the Gaussian distribution is employed to balance the utility and privacy and does not consider the FL settings parameters. The Gaussian levels are set to σ=0.1, σ=0.3, and σ=0.6. For instance, increasing Gaussian noise leads to poor performance, as depicted

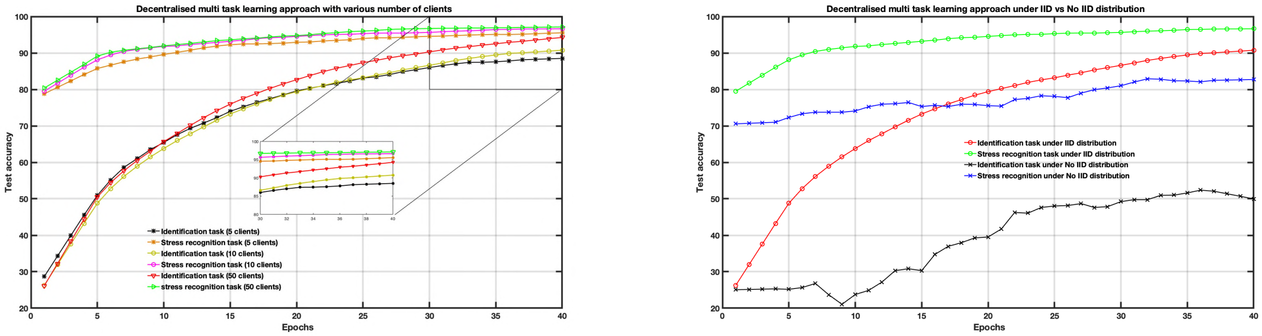

Fig. 3: On the left, the impact of the user participant size on the MFL performance and, on the right, the impact of the IID vs. NO IID distribution on the MFL performance.

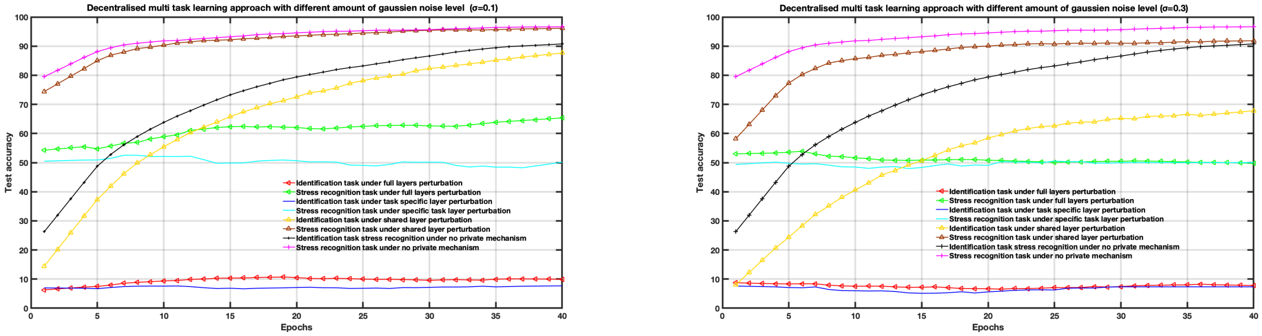

Fig. 4: On the left, the results of the MFL approach under Gaussian noise level (sigma=0.1) and, on the right, the results of the MFL approach under Gaussian noise level (sigma=0.3).

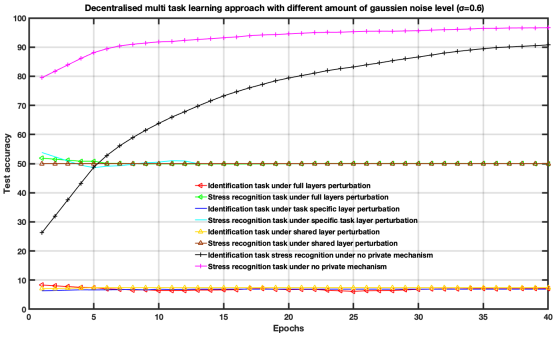

Fig. 5: The results of the MFL approach under Gaussian noise level (sigma=0.6)

in Figure 7.

To examine the DP impact on the utility-privacy tradeoff, we assessed the performance of MFL with a DP mechanism under the aforementioned scenarios. The DP budgets are set as follows ϵ=5, ϵ=15, ϵ=15, ϵ=50 for this experiment. As depicted in Figures 6 and 7 and compared to the baseline scenario, i.e., no private mechanism, adding DP noise into both share and specific layers provides better results for utility performance; however, in terms of privacy, the perturbing specific task layer scheme provides better results than the perturbing shared layer. Results show that FL with perturbing all layers slows up the convergence compared to others, although it provides better privacy (i.e., decreasing identification accuracy).

Intuitively, our results demonstrate that adding noise to upper layers (identity recognition layers) effectively achieves a better privacy-utility tradeoff. This advantage comes at the expense of a formal quantification of the relationship between learning features, i.e., what we aim to share, and private variables, i.e., what we aim to protect, which is rarely available in practice.

We also evaluate the impact of distribution type on the proposed framework performance. The results demonstrate that adding Laplace Noise in our local training model can achieve stress recognition accuracy more than the Gaussian noise types (see Figure 6); however, it maintains the identity recognition task performance.

Nevertheless, employing the Gaussian mechanism (i.e., Gaussian 2) increases the privacy level of the local training model because increasing the number of global iterations will also negatively affect its global convergence performance, i.e., a larger T would increase the noise level variance, dramatically

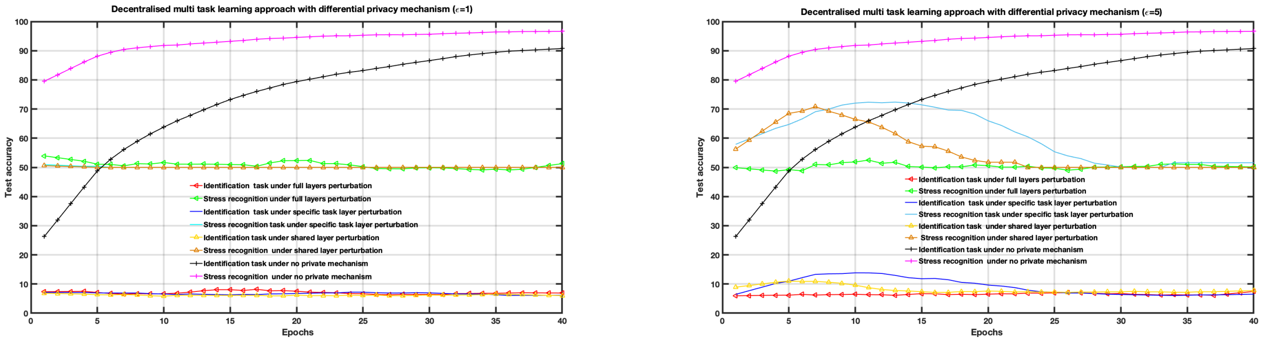

Fig. 6: On the left, the performance of the MFL approach under the DP mechanism (epsilon =1). the performance of the MFL approach under the DP mechanism (epsilon =5) is shown on the right.

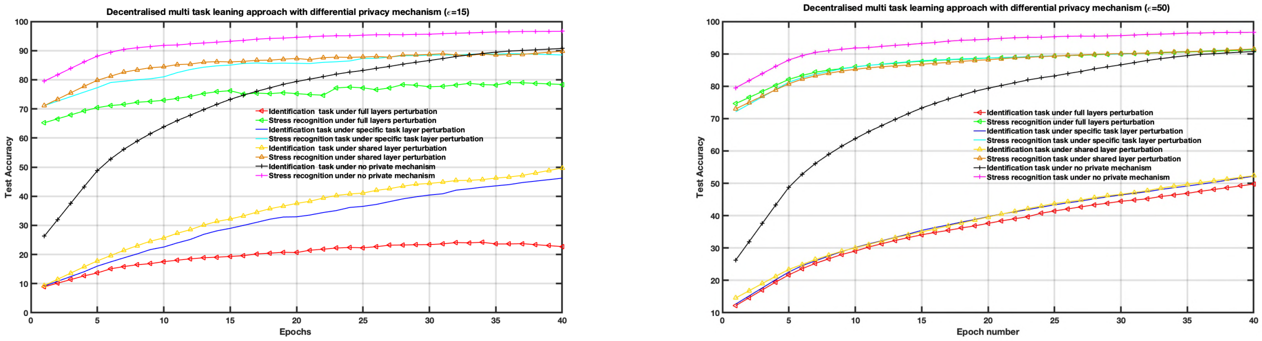

Fig. 7: On the left, the performance of the MFL approach under the DP mechanism (epsilon=15). The performance of the MFL approach under the DP mechanism (epsilon=50) is shown on the right.

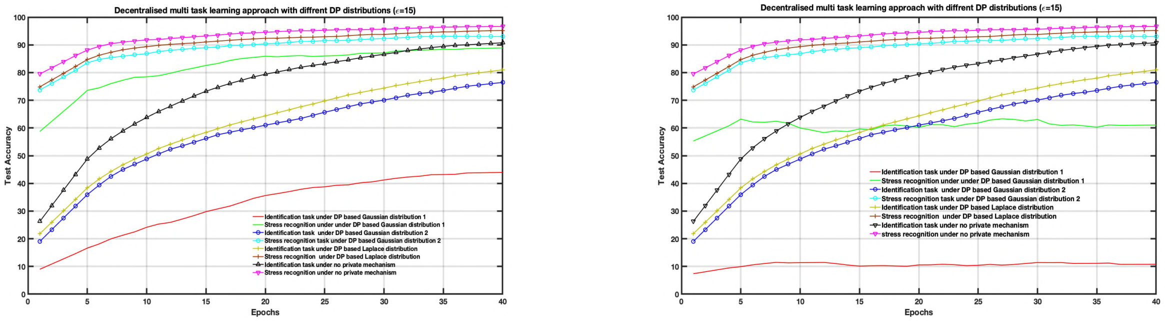

Fig. 8: On the left, the evaluation results of the MFL approach with three DP distributions (specific layer task perturbation), and on the right, the evaluation results of the MFL approach with three DP distributions (Full perturbation).

decreasing the global accuracy model (see Equation 13). In addition, we have found that a larger K contributes to avoiding the vanishing local SGD gradient problem; however, a larger N leads to a scale-down in the variance of noise level injected into the model parameters and fools the SGD training inference.

Furthermore, compared to the FL approach results, the performance of the multi-task FL approach is more consistent. When DP is also used in FL, our experiments suggest that it provides an encouraged result even on lower budget values (i.e., increasing privacy requirements).

V. Conclusion

In this work, we formulate a personalized multi-task federated model framework with differential privacy for a stress recognition system. To satisfy the tradeoff between utility and privacy, We employ a user-level DP mechanism by injecting an amount of noise into personalized layers for perturbing identity while preserving task-specific utility. Our results will guide researchers on the DP privacy-accuracy trade-off for selecting appropriate parameters and distributions according to the tradeoff of utility privacy. Currently, new gradient-based unsupervised adversarial attackers are attacking deep neural classification models to infer the privacy of distributed training gradient. In this case, we aim to provide additional experiments with the federated differentially private generative adversarial networks that can provide better privacy protection and data diversity for widespread applications of physiological computing systems.

VI. Ethical Impact Statement

In order to improve the state of the art in the affective computing field, several new architectures have been tested. MTL is one of the promising ones. However, it can also reveal privacy-sensitive identity information. In practice, it may create privacy issues which can have severe ethical impacts. We tried to alleviate privacy concerns while developing robust systems by using MTL. Having said that, the proposed study has not been tested with a population with sociocultural differences yet.

References

[1] L. C. De Silva, T. Miyasato, and R. Nakatsu, "Facial emotion recognition using multi-modal information," in *Proceedings of ICICS, International Conference on Information, Communications and Signal Processing. Theme: Trends in Information Systems Engineering and Wireless Multimedia Communications (Cat.*, vol. 1. IEEE, 1997, pp. 397–401.

[2] B. Schuller, G. Rigoll, and M. Lang, "Hidden Markov Model-based speech emotion recognition," in *2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03).*, vol. 2, 2003, pp. II–1.

[3] "IDC. (March 10, 2020). Wearables unit shipments worldwide by vendor from 2014 to 2019 (in millions) [graph]." https://www.statista.com/statistics/515634/wearables-shipments-worldwide-by-vendor/, 2020, Accessed at February 25, 2023.

[4] C. Zhang, X. Hu, Y. Xie, M. Gong, and B. Yu, "A privacy-preserving multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition," *Frontiers in neurorobotics*, vol. 13, p. 112, 2020.

[5] Y. Zhang and Q. Yang, "A survey on multi-task learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, no. 12, pp. 5586–5609, 2022.

[6] W. Dai, S. Cahyawijaya, Y. Bang, and P. Fung, "Weakly-supervised multi-task learning for multimodal affect recognition," *arXiv preprint arXiv:2104.11560*, 2021.

[7] M. Abadi, A. Chu, I. Goodfellow, H. B. McMahan, I. Mironov, K. Talwar, and L. Zhang, "Deep learning with differential privacy," in *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security*, 2016, pp. 308–318.

[8] S. Chen, Q. Jin, J. Zhao, and S. Wang, "Multimodal multi-task learning for dimensional and continuous emotion recognition," in *Proceedings of the 7th Annual Workshop on Audio/Visual Emotion Challenge*, ser. AVEC '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 19–26.

[9] D. V. Sang, L. T. B. Cuong, and V. Van Thieu, "Multi-task learning for smile detection, emotion recognition and gender classification," in *Proceedings of the 8th International Symposium on Information and Communication Technology*, ser. SoICT '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 340–347.

[10] Y. Zhao, J. Chen, D. Wu, J. Teng, and S. Yu, "Multi-task network anomaly detection using federated learning," in *Proceedings of the 10th International Symposium on Information and Communication Technology*, ser. SoICT '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 273–279.

[11] V. Smith, C.-K. Chiang, M. Sanjabi, and A. S. Talwalkar, "Federated mtl," *Advances in neural information processing systems*, vol. 30, 2017.

[12] K. Somandepalli, H. Qi, B. Eoff, A. Cowen, K. Audhkhasi, J. Belanich, and B. Jou, "Federated learning for affective computing tasks," in *2022 10th International Conference on Affective Computing and Intelligent Interaction (ACII)*, 2022, pp. 1–8.

[13] D. Shome and T. Kar, "Fedaffect: Few-shot federated learning for facial expression recognition," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 4168–4175.

[14] P. Chhikara, P. Singh, R. Tekchandani, N. Kumar, and M. Guizani, "Federated learning meets human emotions: A decentralized framework for human–computer interaction for iot applications," *IEEE Internet of Things Journal*, vol. 8, no. 8, pp. 6949–6962, 2020.

[15] T. Feng, R. Peri, and S. Narayanan, "User-level differential privacy against attribute inference attack of speech emotion recognition in federated learning," *arXiv preprint arXiv:2204.02500*, 2022.

[16] Y. S. Can and C. Ersoy, "Privacy-preserving federated deep learning for wearable iot-based biomedical monitoring," *ACM Transactions on Internet Technology (TOIT)*, vol. 21, no. 1, pp. 1–17, 2021.

[17] A. Nandi and F. Xhafa, "A federated learning method for real-time emotion state classification from multi-modal streaming," *Methods*, vol. 204, pp. 340–347, 2022.

[18] K. Somandepalli, H. Qi, B. Eoff, A. Cowen, K. Audhkhasi, J. Belanich, and B. Jou, "Federated learning for affective computing tasks," in *2022 10th International Conference on Affective Computing and Intelligent Interaction (ACII).* IEEE, 2022, pp. 1–8.

[19] S. Latif, S. Khalifa, R. Rana, and R. Jurdak, "Federated learning for speech emotion recognition applications," in *2020 19th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN).* IEEE, 2020, pp. 341–342.

[20] C. Dwork, A. Roth *et al.*, "The algorithmic foundations of differential privacy," *Foundations and Trends® in Theoretical Computer Science*, vol. 9, no. 3–4, pp. 211–407, 2014.

[21] K. Wei, J. Li, M. Ding, C. Ma, H. Su, B. Zhang, and H. V. Poor, "User-level privacy-preserving federated learning: Analysis and performance optimization," *IEEE Transactions on Mobile Computing*, 2021.

[22] P. Schmidt, A. Reiss, R. Duerichen, C. Marberger, and K. Van Laerhoven, "Introducing wesad, a multimodal dataset for wearable stress and affect detection," in *Proceedings of the 20th ACM international conference on multimodal interaction*, 2018, pp. 400–408.

[23] B. Cinaz, B. Arnrich, R. La Marca, and G. Tröster, "Monitoring of mental workload levels during an everyday life office-work scenario," *Personal and ubiquitous computing*, vol. 17, pp. 229–239, 2013.

[24] D. C. Toledo-Pérez, J. Rodríguez-Reséndiz, R. A. Gómez-Loenzo, and J. Jauregui-Correa, "Svm-based emg signal classification techniques: A review," *Applied Sciences*, vol. 9, no. 20, p. 4402, 2019.

[25] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Artificial intelligence and statistics*. PMLR, 2017, pp. 1273–1282.

[26] M. G. Arivazhagan, V. Aggarwal, A. K. Singh, and S. Choudhary, "Federated learning with personalization layers," *arXiv preprint arXiv:1912.00818*, 2019.

[27] T. Liu, B. Di, B. Wang, and L. Song, "Loss-privacy tradeoff in federated edge learning," *IEEE Journal of Selected Topics in Signal Processing*, vol. 16, no. 3, pp. 546–558, 2022.