Balancing Between Privacy and Utility for Affect Recognition using Multi Task Learning in Differential Privacy Added Federated Learning Settings

Mohamed Benouis, Elisabeth Andre, Yekta CAN

Submitted to: JMIR Mental Health
on: April 28, 2024

Table of Contents

Balancing Between Privacy and Utility for Affect Recognition using Multi Task Learning in Differential Privacy Added Federated Learning Settings

Mohamed Benouis[1]; Elisabeth Andre[1]; Yekta CAN[1]

[1]Chair for Human Centered AI, Faculty of Applied Computer Science Augsburg University Augsburg DE

Corresponding Author:
Yekta CAN
Chair for Human Centered AI, Faculty of Applied Computer Science
Augsburg University
Universitätsstraße 6a
Augsburg
DE

Abstract

Background: The rise of wearable sensors marks a pivotal development in the era of affective computing. These sensors, gaining increasing popularity, hold the potential to revolutionize our understanding of human stress. A fundamental aspect within this domain is the ability to discern perceived stress through these unobtrusive devices.

Objective: This study aims to enhance the performance of emotion recognition utilizing multi-task learning, a technique extensively explored across various machine learning tasks, including affective computing. By leveraging the shared information among related tasks, we seek to augment the accuracy of emotion recognition while confronting the privacy threats inherent in the physiological data captured by these sensors.

Methods: To address the privacy concerns associated with the sensitive data collected by empathetic sensors, we propose a novel framework that integrates differential privacy and federated learning approaches with multi-task learning. This framework is designed to efficiently identify the user's mental stress while safeguarding their private identity information. Through this approach, we aim to enhance the performance of emotion recognition tasks while preserving user privacy.

Results: Extensive evaluations of our framework are conducted using two prominent public datasets. The results demonstrate a significant improvement in emotion recognition accuracy, achieving an impressive rate of 90%. Furthermore, our approach effectively mitigates privacy risks, as evidenced by limiting re-identification accuracies to 47%.

Conclusions: In conclusion, our study presents a promising approach to advancing emotion recognition capabilities while addressing privacy concerns in the context of empathetic sensors. By integrating multi-task learning with differential privacy and federated learning, we have demonstrated the potential to achieve high levels of accuracy in emotion recognition while ensuring the protection of user privacy. This research contributes to the ongoing efforts to harness the power of affective computing in a responsible and ethical manner.

(JMIR Preprints 28/04/2024:60003)
DOI: https://doi.org/10.2196/preprints.60003

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Original Paper

Balancing Between Privacy and Utility for Affect Recognition using Multi Task Learning in Differential Privacy Added Federated Learning Settings

Mohamed Benouis, Elisabeth André, Yekta Said Can

Chair for Human Centered AI, Faculty of Applied Computer Science, University of Augsburg, Augsburg, GERMANY

Abstract

Background: The rise of wearable sensors marks a pivotal development in the era of affective computing. These sensors, gaining increasing popularity, hold the potential to revolutionize our understanding of human stress. A fundamental aspect within this domain is the ability to discern perceived stress through these unobtrusive devices.

Objective: This study aims to enhance the performance of emotion recognition utilizing multi-task learning, a technique extensively explored across various machine learning tasks, including affective computing. By leveraging the shared information among related tasks, we seek to augment the accuracy of emotion recognition while confronting the privacy threats inherent in the physiological data captured by these sensors.

Methods: To address the privacy concerns associated with the sensitive data collected by empathetic sensors, we propose a novel framework that integrates differential privacy and federated learning approaches with multi-task learning. This framework is designed to efficiently identify the user's mental stress while safeguarding their private identity information. Through this approach, we aim to enhance the performance of emotion recognition tasks while preserving user privacy.

Results: Extensive evaluations of our framework are conducted using two prominent public datasets. The results demonstrate a significant improvement in emotion recognition accuracy, achieving an impressive rate of 90%. Furthermore, our approach effectively mitigates privacy risks, as evidenced by limiting re-identification accuracies to 47%.

Conclusions: In conclusion, our study presents a promising approach to advancing emotion recognition capabilities while addressing privacy concerns in the context of empathetic sensors. By integrating multi-task learning with differential privacy and federated learning, we have demonstrated the potential to achieve high levels of accuracy in emotion recognition while ensuring the protection of user privacy. This research contributes to the ongoing efforts to harness the power of affective computing in a responsible and ethical manner.

Keywords: privacy preservation, multitask learning, federated learning, differential privacy, physiological signals, affective computing.

Introduction

Imagine awakening in the morning to find yourself grappling with stress, perhaps stemming from a disagreement with a friend, a pressing financial concern, or an unresolved work-related matter. In this scenario, an advanced virtual assistant, akin to Siri or Alexa, seamlessly detects your stress level through the analysis of physiological signals captured by wearable technology. Leveraging this data, the virtual assistant dynamically adjusts its language and tone to suit your mood. Moreover, it proactively recommends personalized relaxation techniques, drawing from past

successful interventions such as yoga, mindfulness practices, your favorite music, and uplifting videos. Such a sophisticated stress recognition system holds promise for enhancing the well-being and emotional resilience of individuals. Although most of the studies use facial expressions [1] and speech [2] for recognizing stress and emotions (i.e. affects as an umbrella term) physiology-based methodologies also emerged as an alternative since they can offer promise for seamless, continuous monitoring within everyday contexts. Notably, the year 2020 witnessed the sale of over 330 million smartwatches, fitness trackers, and analogous wearables capable of gathering quantitative physiological data [3], thereby positioning them as auspicious tools for affect recognition owing to their unobtrusive data collection capabilities.

Multi-task learning (MTL) is a method proposed to tackle training multiple related tasks at the same time. It works by sharing knowledge between these tasks to make each model perform better [4]. Essentially, MTL acts like a behind-the-scenes helper, enhancing the ability of machine learning (ML) models to generalize different types of data [5]. This technique has been particularly useful in improving ML models across various fields, including affective computing [6]. For instance, if we want to improve how a system recognizes emotions, we can also train it to recognize gender [7]. By doing this, the system can learn from both tasks simultaneously, making it more effective in recognizing emotions and enhancing performance. However, there's a concern when third parties gain access to additional information, like gender or age, which could be misused for targeted advertising or even discrimination in things like job opportunities [8]. Moreover, if sensitive information like biometric data gets leaked, it could lead to cybercrimes such as identity theft. Therefore, it's crucial for MTL systems to incorporate privacy-preserving methods.

Researchers have introduced Federated Learning (FL) as a way to address privacy issues. FL allows users to train their data locally while sharing only the trained model parameters, rather than the original data itself. It's like bringing the tools to work on the data, instead of moving the data around [9]. This method complies with privacy regulations such as the EU General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA). FL has shown promise in safeguarding user privacy in Internet of Things (IoT) networks.

Although FL is a significant step forward in protecting user privacy, it's not foolproof. There's still a risk of sensitive information being uncovered by reverse engineering the local model parameters. To further enhance privacy, researchers have integrated privacy-preserving techniques into ML for physiological data. One such method is Differential Privacy (DP), which adds random noise to each model in the client or server, disrupting updates and limiting the leakage of sensitive information between nodes [10]. However, it's worth noting that DP can reduce the performance of ML models.

In practice, protecting the user’s privacy without degradation in model utility is still an open problem. In this study, we first 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 order to obtain a robust performance, we separated the two tasks and added noise to only the privacy task which is biometric identity recognition. 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.

Related Works

In order to develop a robust affective computing system that can be used in practical applications, researchers tested various modalities with 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. Chen et al. [10] applied it to audiovisual signals in the AVEC challenge and achieved a significant performance increase compared to baseline ML algorithms by detecting arousal and valence levels simultaneously. Since the human face source can be used for several tasks, such as gender recognition or age estimation, Sang et al. used MTL with CNN for smile detection, emotion recognition, and gender classification [7]. They outperformed the state-of-the-art in selected three benchmark datasets (IMDB and Wiki dataset, GENKI-4K dataset, and FERC-2013 dataset). MTL-based centralized learning for affective computing is beneficial for simultaneously sharing features and learning auxiliary tasks. 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 especially while uploading the features to the central server for both utility and auxiliary tasks for learning.

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. 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 fairly.

FL has been widely applied for facial features for affect recognition. Somandepalli et al. [13] investigated FL for two affective computing tasks: classifying self-reports and perception ratings in audio, video, and text datasets. Using the speech modality, Latif et al. [20] investigated an FL approach in emotion recognition tasks while sharing only the model among clients. They implemented an LSTM classifier and tested it on the IEMOCAP dataset with four emotions: happy, sad, angry, and neutral. Chhikara et al. [15] combined emotion recognition from facial expressions and speech with an FL approach. For the face modality, they employed a combination of CNN and SVM models, whereas for the audio modality, they applied a 2D CNN model to extracted spectrogram images. Their proposed framework has been validated and tested on two datasets, FER2013 for facial emotion recognition and Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS) for speech emotion recognition, respectively.

Table 1. Studies using MTL or Privacy-preserving approaches for various applications.

Study	Application	Signal	MTL	FL	DP	Multimodality
Zhao et al. [11]	Network Anomaly Detection	Network Signals	✓	✓	X	X
Smith et al. [12]	Activity recognition	Acceleration, Gyroscope	✓	✓	X	✓
Somandepalli et al. [13]	Affect recognition	Speech, video, text	X	✓	X	✓
Sang et al. [7]	Affect recognition	Face images	✓	X	X	X
Shome and Kar [14]	Affect recognition	Face images	X	✓	X	X
Chhikara et al. [15]	Affect recognition	Face and speech	X	✓	X	✓
Feng et al. [16]	Affect recognition	Speech	X	✓	✓	X
Can and Ersoy [17]	Affect Recognition	PPG	X	✓	X	X

Nandi and Xhafa [18]	Affect Recognition	EEG, EDA	X	✓	X	✓
This Study	Affect Recognition	PPG, EDA, Acceleration, ST	✓	✓	✓	✓

On the contrary, Federated Learning (FL) has been seldom utilized in the context of affect recognition from physiological signals, as indicated in Table I. Can and Ersoy [17] implemented an FL learning model to forecast perceived stress utilizing physiological data. Each sub-client employs an MLP classifier to locally train its data on the edge, and the sharing of MLP's individual updating parameters is facilitated using the FedAVG algorithm. FL has also been extended to handle multimodal physiological signals. Nandi and Xhafa [18] proposed Fed-ReMECS, an FL framework-based machine learning model for emotion recognition. They applied wavelet feature extraction and a neural network to Electrodermal Activity (EDA) and respiration data from the DEAP dataset to recognize valence arousal levels, validating their approach. These studies demonstrate the successful application of FL without compromising affect recognition performance.

While FL has been introduced to enhance model training in terms of privacy, the privacy vulnerabilities inherent in the stochastic gradient descent (SGD) algorithm remain unresolved. Differential Privacy (DP) mechanisms have primarily been discussed in FL settings, involving the injection of noise into each model client or server to perturb updates and limit gradient leakage shared among nodes (i.e., client and server) [21]. In one of the initial applications, authors introduced a new private training method termed differential private stochastic gradient descent, which reduces local and global gradient information leakage between the client and server. Instead of employing the standard composition theorem to calculate the final distribution of overall noise clients, they utilized a moments accountant metric to adaptively monitor the overall privacy loss. Recognizing that servers are often curious or untrustworthy, Wei et al. [22] proposed a local DP mechanism algorithm by introducing Gaussian noise distribution into user models before uploading them to servers. To address the communication overhead required for optimal convergence upper bound for DP, they introduced a novel approach known as communication rounds discounting (CRD) method, which achieves a better trade-off between the computational complexity of searching and convergence performance. DP has also been leveraged for affect recognition. Feng et al. [16] utilized user-level differential privacy (UDP) to mitigate privacy leaks in FL for speech emotion recognition. Recently, Smith et al. [12] showcased promising performance in preserving privacy via multi-task FL for activity recognition.

When examining the affective computing literature, we find studies using MTL to improve performance and others using FL or DP to preserve privacy. However, there are no studies that leverage MTL for performance enhancement while also preserving privacy in this literature. The overall idea of our approach is to separate the utility and the privacy tasks preventing DP from compromising the performance of utility task (affect recognition) and introducing noise exclusively to the privacy task (biometric identity recognition). Consequently, our work represents the first MTL-based approach to affect recognition using FL and DP to preserve privacy at the same time.

Methods

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

A. Data Description and Feature Extraction

For our experiments, we use two datasets WESAD [23] and VERBIO [33], which have been created for affective state monitoring.

WESAD: In WESAD dataset, 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 [23].

VERBIO: The VERBIO dataset [33] consists of bio- behavioral responses and self-reported data during public speaking presentations, both in front of real-life and virtual audiences. 55 participants conducted 10 distinct presentations, each lasting approximately 5 minutes, across the three segments of the study, spanning four days: PRE (1 session, Day 1), TEST (8 sessions, Days 2-3), and POST (1 session, Day 4). The PRE and POST segments involved real-life audiences, while the TEST segment featured various virtual audiences. To prevent participant exhaustion, the TEST segment was divided into two days, each comprising four sessions. In total, the study yielded 10,800 minutes of acoustic and physiological data from 82 real and 216 virtual reality (VR) presentations. In total, 35174 segments were created. We solely utilized physiological data collected using the Empatica E4 device, possessing identical sampling rates and modalities as those employed in the WESAD dataset.

Preprocessing: Each modality of the signal is segmented using 700 sample windows size with 50% overlap, as suggested in the literature [24]. To maximize the correlation among inter-subjects and minimize among subjects, these segments were further processed for extracting features [17] 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 No IID [8].
2) For the local training process, there is only one iteration for SGD local training for each client. In particular, w is the local model parameter [8], given by:

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

3) The local data of different clients cannot be communicated and only the models can be

shared.

4) Following the Federated Averaging algorithm (FedAvg) [8], the server employs a global averaging approach to aggregate all local training models to compute the final global model. Formally [8], the server aggregates the weights sent from the K clients as:

$$w = \sum_{U=1}^{K} p_i w_U^{D_i} \quad (2)$$

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, $D = U D_i$ the whole distributed dataset, and $P_U = |D_i| / ¿ D \vee ¿$.

5) The global training epoch is set to M rounds (aggregations). The server solves the optimization problem [8]

$$W^{¿} = arg\,min \sum_{U=1}^{M} P_U F_U \left(w_U, D_i \right) \quad (3)$$

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, bio- metric template user, etc. In such cases, the Multitask Federated Learning (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. We also compared adding noise to whole model and task specific last layers in addition to the shared layers and demonstrated the performance of these three different approaches (see Figure 1). Motivated by personalized FL [26], our work focuses on client-level privacy, which aims at 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 layers. 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:

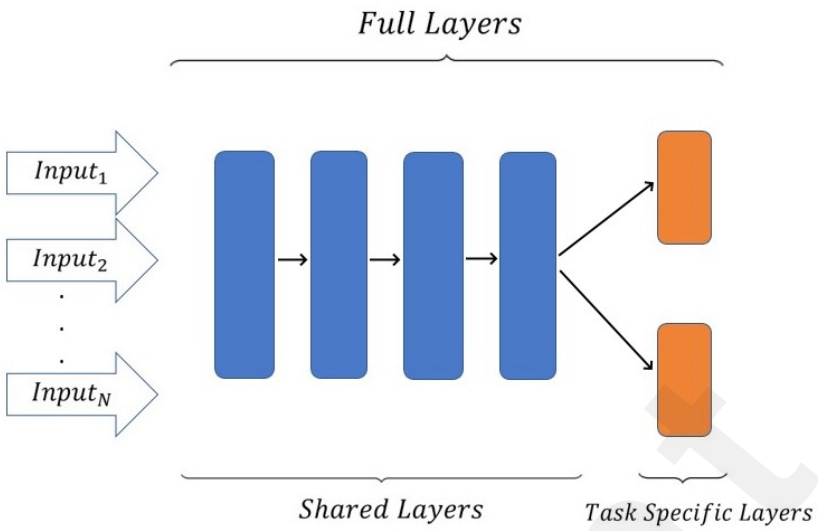

Figure 1: Three different scenarios for adding noise with DP. Adding noise to only shared layers, task-specific layers and full layers.

1) An additive Gaussian noise η∼N 0, σ to each weight. This operation can be mathematically described as follows:

$$w_{t+1} = w_t + \eta \quad (4)$$

2) A set of noise distributions can be sampled from the DP mechanism (DP). A randomized mechanism M on the training set with domain X and range R satisfies (ϵ, δ) − DP for two small positive numbers and if the following inequality holds [21]:

$$Pr[M(x) \in S] \leq e^{\epsilon} Pr[M(x') \in S] + \delta \quad (5)$$

where x and $x' \in X$ 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:

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

where N is scaling noise sampled from specific distribution. In this work, we chose Laplace and Gaussian mechanisms [21] that employ L1 and L2 norm sensitivity, respectively. The sensitivity function can be expressed as:

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

And scaling noise can be computed as:

$$\sigma = \Delta f / \varepsilon \quad (8)$$

Output perturbation satisfies $(\epsilon, \delta) - DP$ when we properly select the value scaling noise. Thus, it sampled from Laplace and Gaussian distributions [21] as:

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

$$M_{Gaussian}(x, f, \varepsilon, \delta) = f(x) + N(\mu = 0, \sigma^2) \tag{10}$$

The gradient information leakage can be reduced by applying gradient thresholding or clipping algorithm. As explained in [9], 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.

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 σ^2 and DP parameter initializations (ϵ_i, δ_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
    $\mathbf{w}_i^{(t)} = \mathbf{w}_i^{(t)} / \max\left(1, \frac{\left\| \mathbf{w}_i^{(t)} \right\|}{C}\right)$
      ▷ Perturb selected layers (full, shared, task)
    ▷ Add with DP or Gaussian noise, i.e., $\eta_{\sim} \mathcal{N}(0, \sigma_l^2)$
    $\tilde{\mathbf{w}}_i^{(t)} = \mathbf{w}_i^{(t)} + \mathbf{n}_i^{(t)}$
  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 : $\widetilde{\mathbf{w}}^{(T)}$

Experimental Results

Three scenarios are created to tackle the aforementioned challenges with the DP learning approaches: centralized, decentralized FL, and decentralized FL with DP. Their performances are evaluated on WESAD and VERBIO datasets 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 train a multi-task deep learning model for handling these tasks simultaneously. In addition, for better training and to avoid overfitting, an Early Stopping regularization technique is used as gradient descent. 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).

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 ($Loss_T$).

Table 2. Parameters of the Multitask 1D CNN based Architecture.

Layer	Type	Parameters
Input	Input	Features Size
Conv1D	Convolution	Input=1, Output=20, K=8, Stride=1, Padding
ReLu	Activation Function	ReLu
Pooling	Pooling	Stride=2, Max Pooling
Conv	Convolution Layer	Input=20, Output=40, K=8, Stride=1, Padding
Relu	Activation Function	ReLu
Pooling	Pooling	Stride=2, Max Pooling
Conv	Convolution Layer	Input=40, Output=60, K=8, Stride=1, Padding
Relu	Activation Function	ReLu
Pooling	Pooling	Stride=2, Max Pooling
Fully Connected 1	Fully Connected Layer	Input=360, Output=100
Fully Connected 2	Fully Connected Layer	Input=100, Output=300
Linear 1	Input:100 Output:2	Output=2 affect classes, activation function:linear
Linear 2	Input:300 Output:N	Output=N identity classes, activation function:linear

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 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, employing a 1D-CNN multi-task model for stress and identity recognition tasks on two datasets: WESAD and VERBIO. After training, the output layers are used to infer the stress level and the identity of the user, resulting in quite similar average scores of 99%. The results reveal a potential information leakage in this case wherein model accuracy is preserved at the expense of user privacy. As a result, this approach fails to ensure the users' privacy since their data is transmitted to the server for training purposes.

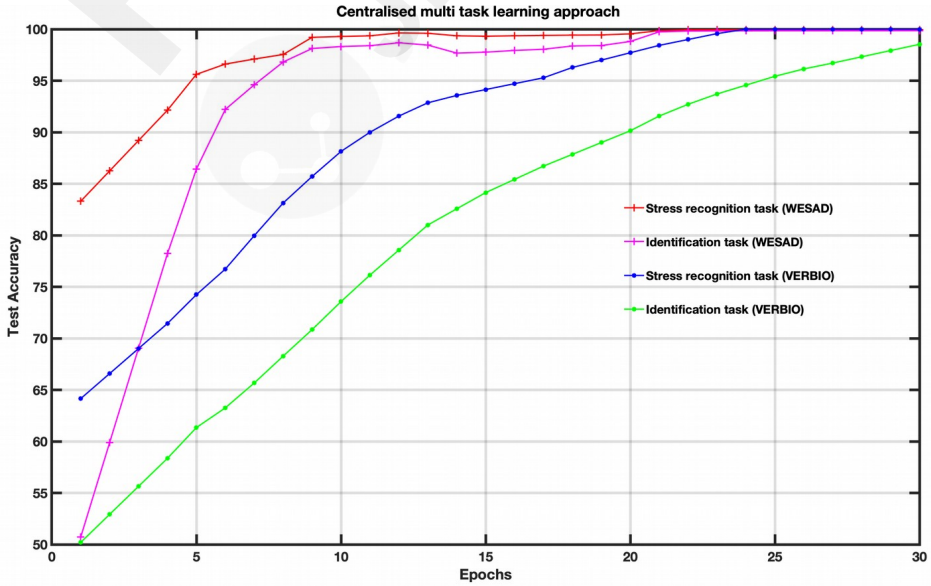

Figure 2: Stress vs. identity recognition performance of multi-task CL using two different datasets.

B. Multitask Decentralized FL Approach (MFL)

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 average accuracy result achieved is 97% and 95% for stress mood recognition and 93% and 90% for user identification on WESAD and VERBIO respectively.

To examine the effect of client participation within the multi-task FL model, we tested different numbers of clients, i.e., K=5, K=10, and K=20. As reported in [25] 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 for both tasks: stress recognition and subject identification (see Figures 4 and 5). We 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.

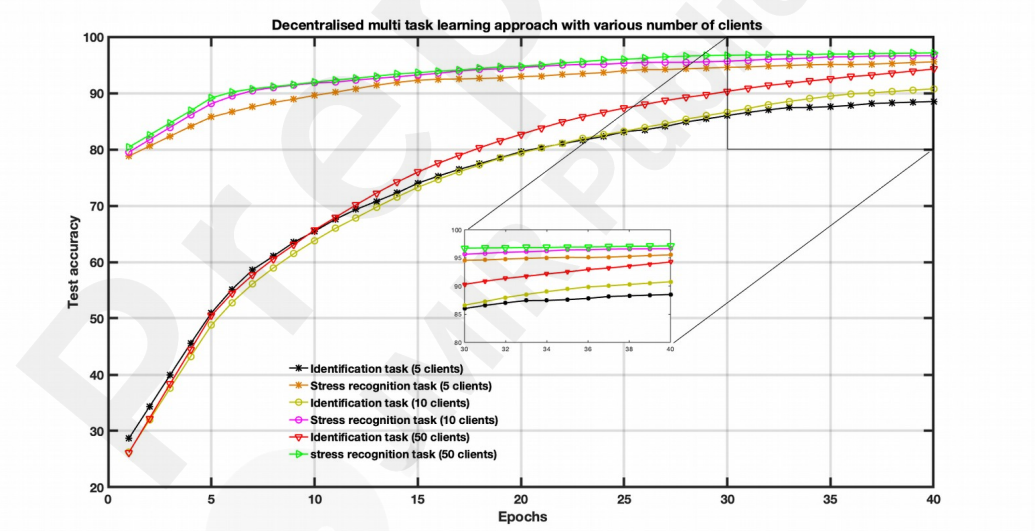

Figure 3: The impact of the user participant size on the MFL performance (WESAD)

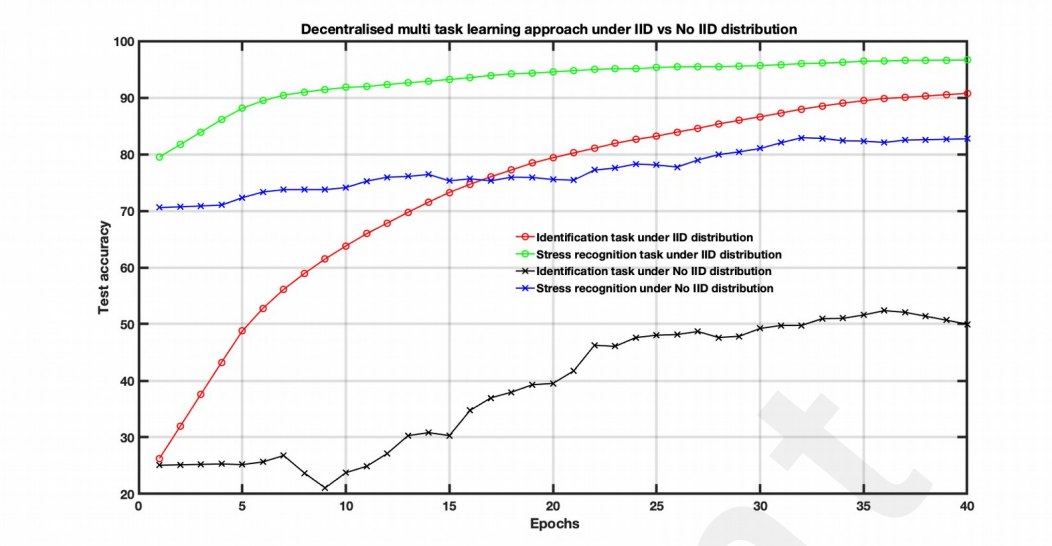

Figure 4: The impact of the IID vs. NO IID dist. on the MFL performance (WESAD).

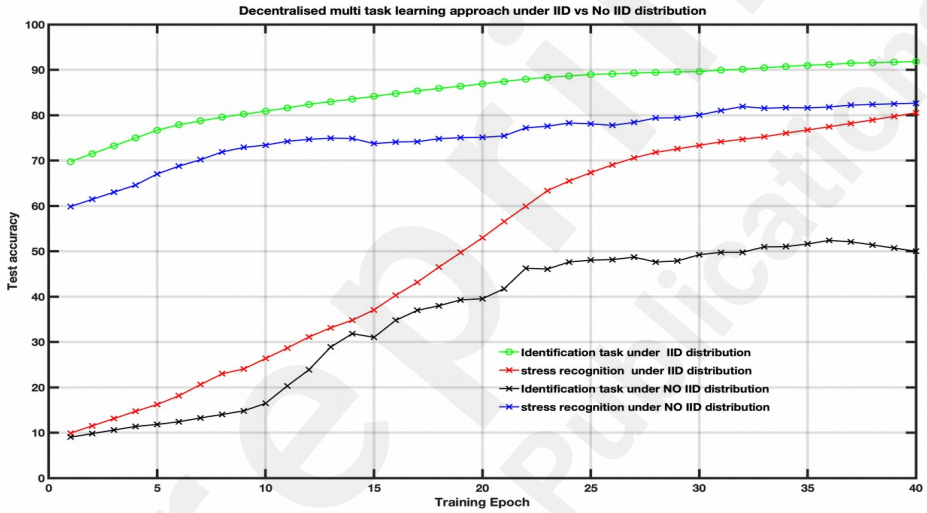

Figure 5: The impact of the IID vs. NO IID dist. on the MFL performance (VERBIO)

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., $\eta \sim N(0,\sigma^2)$. 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 [9] is computed as:

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

a) Gaussian distribution, two distributions are given by [9] and [22] respectively.

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

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

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 6, 7 and 8 show the accuracy comparison when adding Gaussian distribution levels into local training according to the three scenarios on the WESAD dataset. 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%.

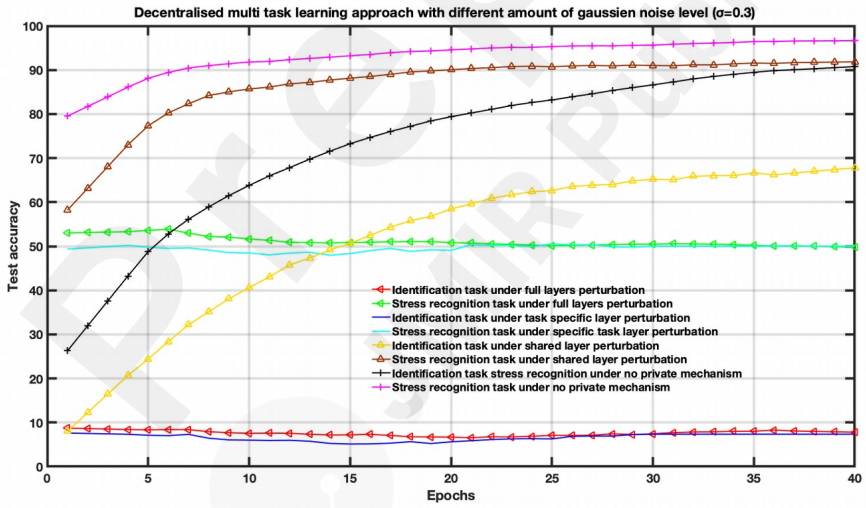

Figure 6: MFL approach results under Gaussian noise level (sigma=0.1)

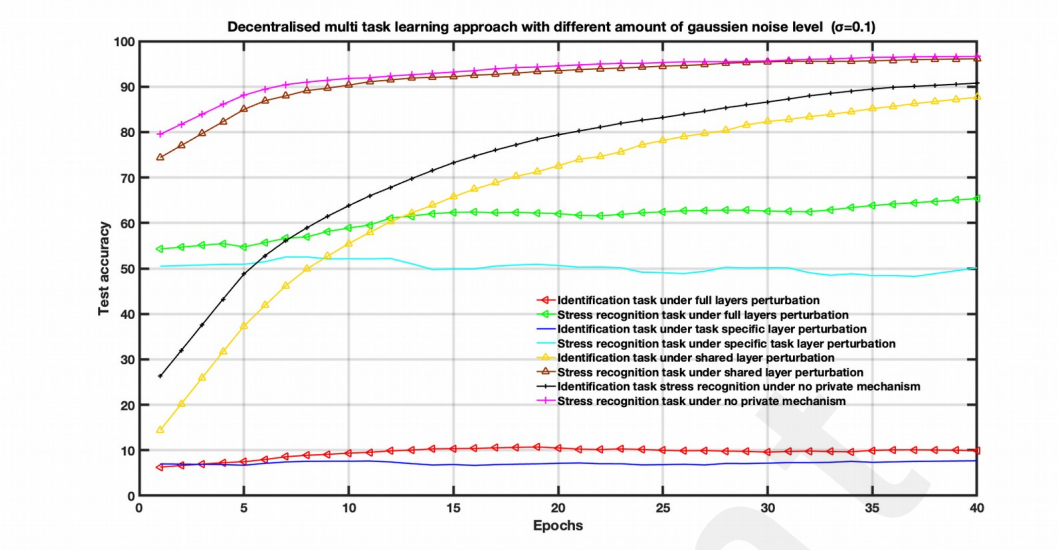

Figure 7: MFL approach results under Gaussian noise level (sigma=0.3).

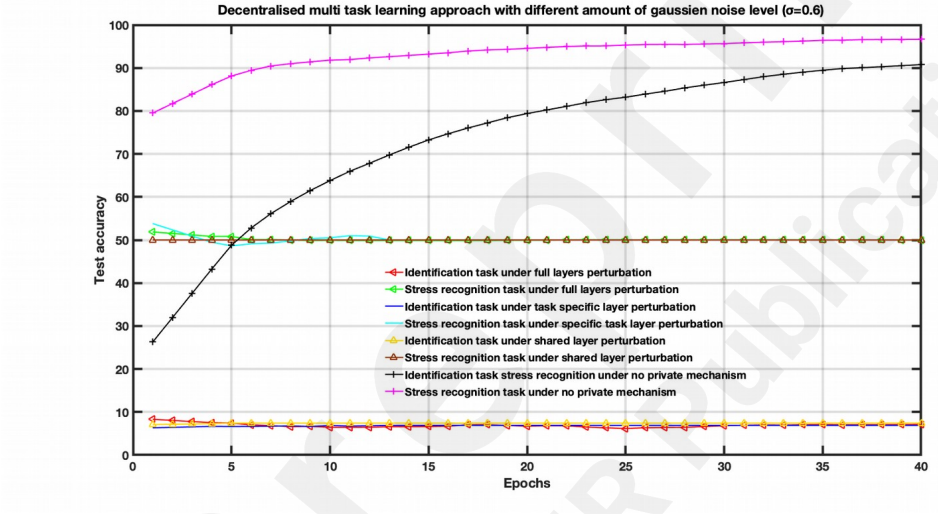

Figure 8: MFL approach results under Gaussian noise level (sigma=0.6)

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 in Figure 8.

To examine the DP impact on the utility-privacy tradeoff, we assessed the performance of MFL with a DP mechanism in WESAD and VERBIO datasets under the aforementioned scenarios. The DP budgets are set as follows ϵ=5, ϵ=15, ϵ=50 for this experiment. As depicted in Figures 9-16 compared to the baseline scenario, i.e., no privacy enhancing mechanism, adding DP noise into both shared 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).

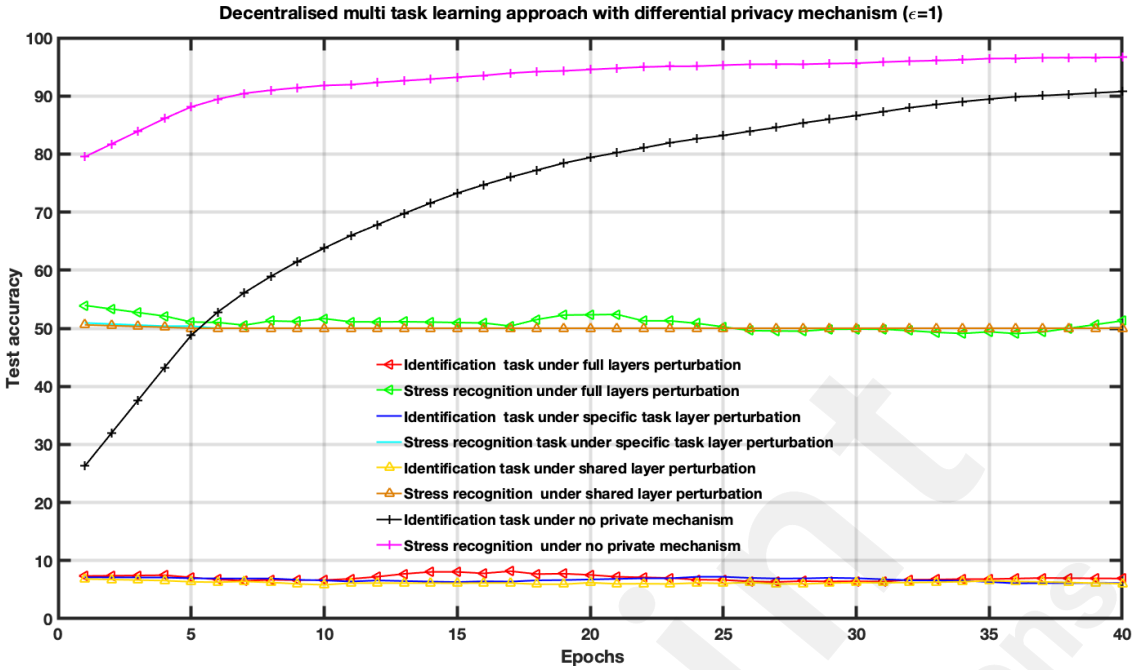

Figure 9: The performance of the MFL approach under the DP mechanism (WESAD) (epsilon =1).

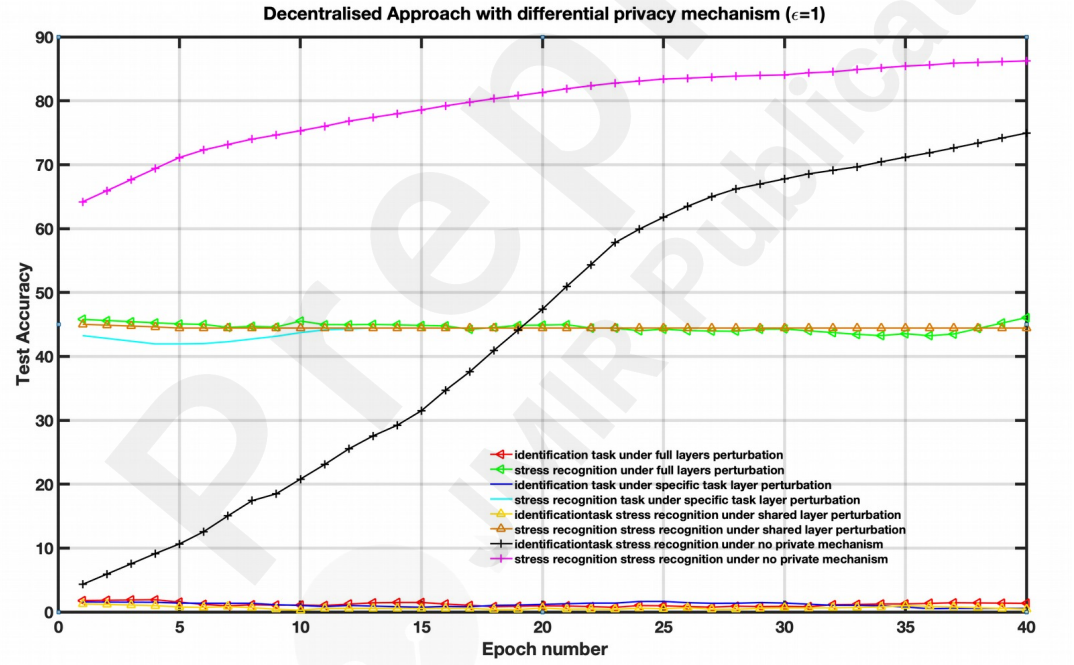

Figure 10: The performance of the MFL approach under the DP mechanism (VERBIO) (epsilon =1).

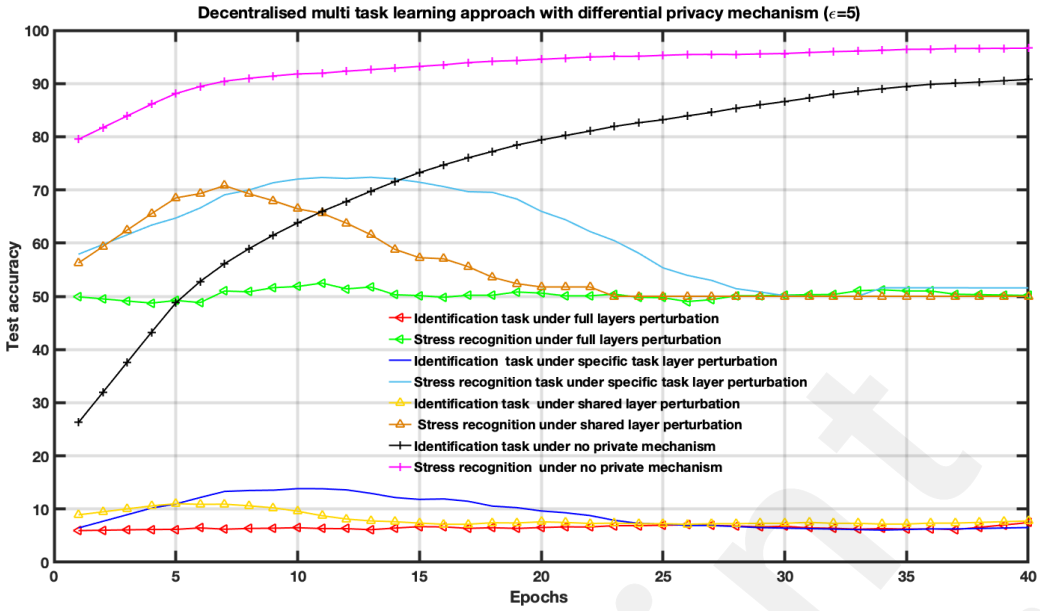

Figure 11: the performance of the MFL approach under the DP mechanism (WESAD) (epsilon =5)

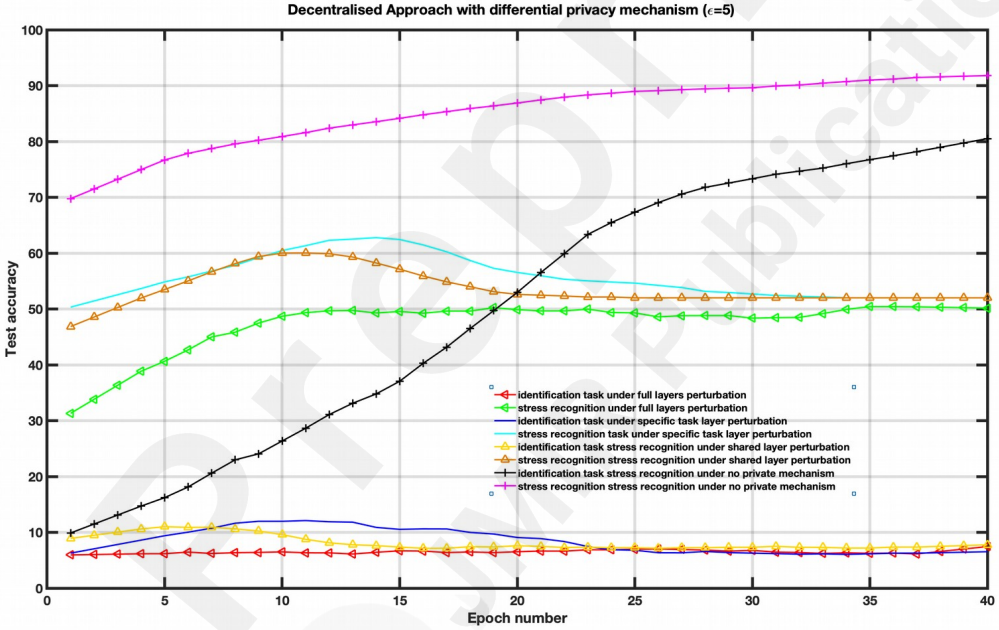

Figure 12: The performance of the MFL approach under the DP mechanism (VERBIO) (epsilon =5).

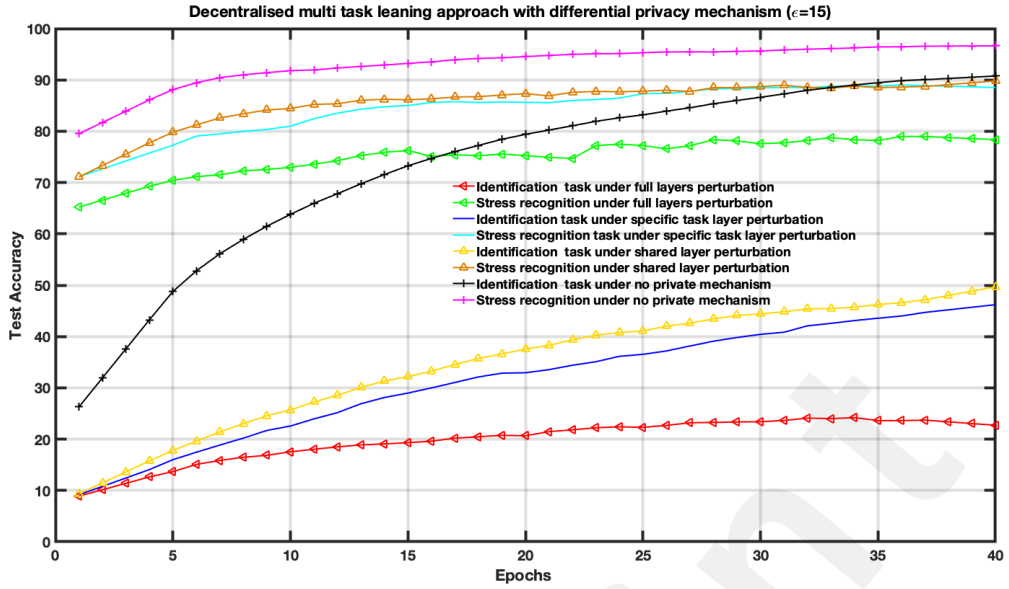

Figure 13: the performance of the MFL approach under the DP mechanism (WESAD) (epsilon=15).

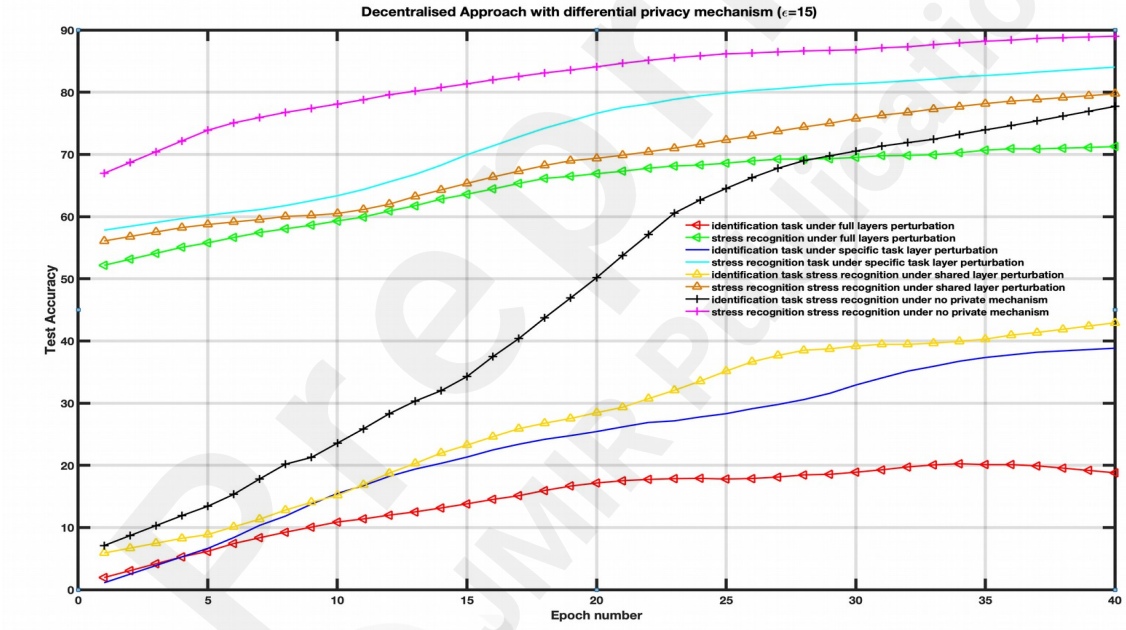

Figure 14: The performance of the MFL approach under the DP mechanism (VERBIO) (epsilon =15).

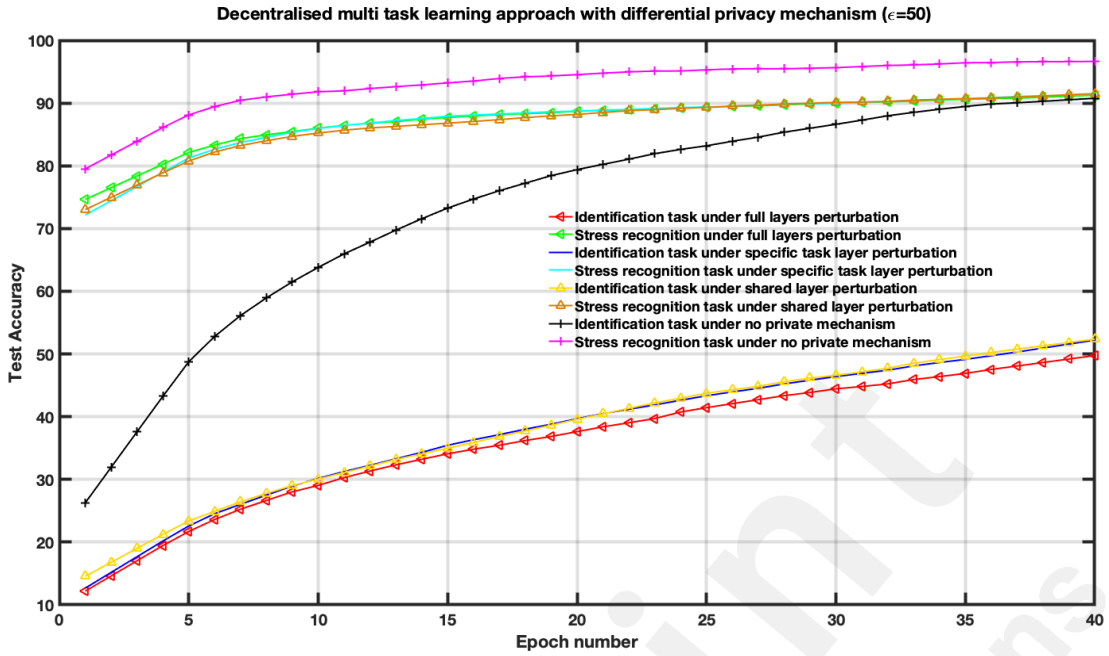

Figure 15: The performance of the MFL approach under the DP mechanism (WESAD) (epsilon=50).

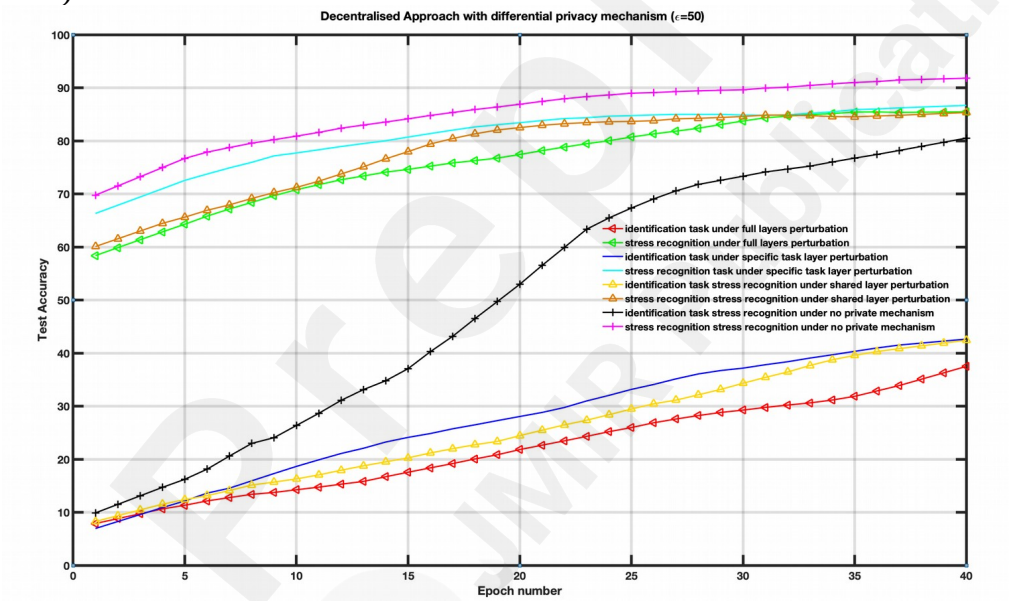

Figure 16: The performance of the MFL approach under the DP mechanism (VERBIO) (epsilon =50).

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 noise distribution type on the proposed framework performance. The results from the WESAD dataset demonstrate that adding Laplace Noise in our local training model preserves the stress recognition accuracy better than the Gaussian noise types (see Figures 17 and 18). However, it also maintains the identity recognition task performance.

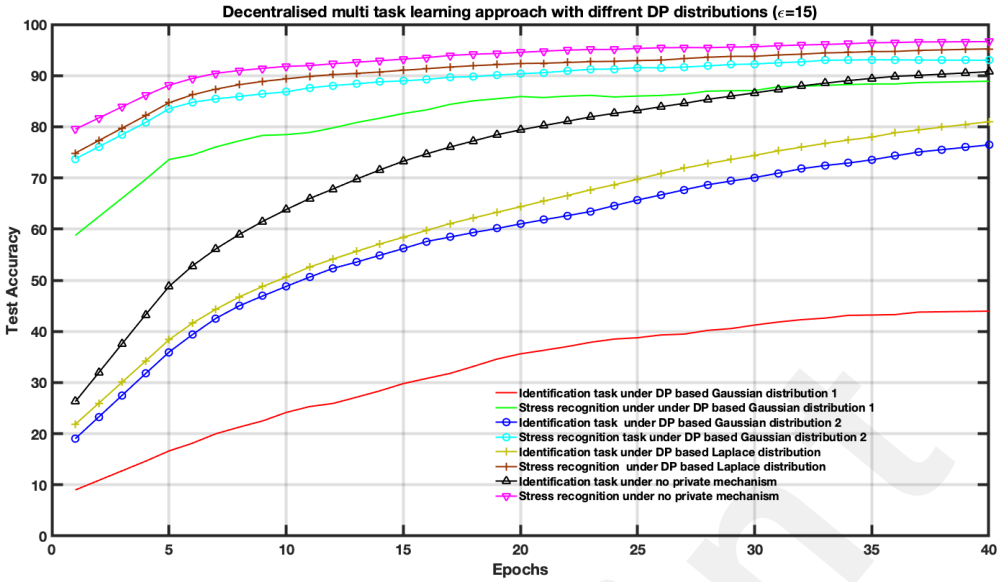

Figure 17: The evaluation results of the MFL approach with three DP distributions (specific layer task perturbation).

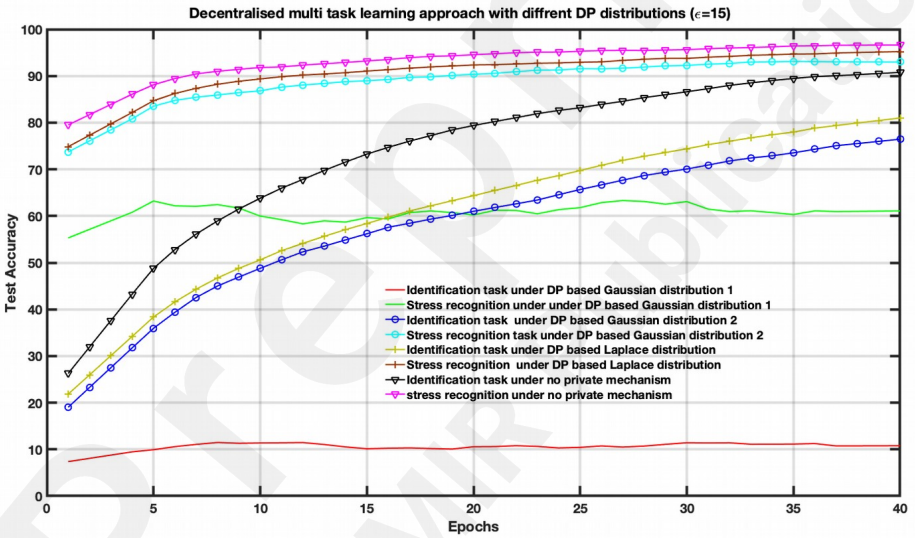

Figure 18: The evaluation results of the MFL approach with three DP distributions (Full perturbation).

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 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 injected noise level to model parameters and fools the SGD training inference.

When DP is also used in FL settings, our experiments suggest that it achieves encouraging performance even on lower budget values (i.e., increasing privacy requirements). The compromise of physiological data privacy can have significant consequences for an individual's life. It provides an opening for data attackers and puts them at danger of data breaches, which can result in several threats [27]. These risks include the possibility of revealing a soft biometric (e.g., gender, location, authenticity, etc) and hard biometric (i.e. identity). So far, most of the research in using physiological

modalities in both biometric domains focused on ECG, EEG, EDA and PPG [26, 27]. We found that the model obviously achieves better balance between the stress and identification classification task when multimodal signals are combined and used as an input, compared to the single modality. However, as aimed in our work, it's difficult to provide privacy guarantees since there is no study provides an investigation on which modality can reveal sensitive information about the subject. In this experiment, we provide a comparative study among these employed modalities for both tasks, subject identification and emotion recognition. According to Table 3, we found that the ECG and acceleration modalities always led to an increase in the performance of the identification task. This fact might have contributed to their popularity in biometric field [28]. This observation hints us that discarding model parameters containing sensitive information learned from modalities associated with identity information might help to effectively achieve a good privacy-utility tradeoff to a certain extent, that is, decrease information leakage.

Table 3: The effect of modality on the stress and subject identification performance

	Modalities	Task	Approach		
			Centralized	FL	FL + DP
WESAD	ACC	Stress	91.60	89.90	84.46
		Subject Identification	95.50	94.88	58.19
	BVP	Stress	98.34	96.05	90.97
		Subject Identification	94.96	92.05	54.95
	EDA	Stress	99.00	95.29	91.27
		Subject Identification	86.15	85.89	50.05
	ECG	Stress	97.01	93.37	91.60
		Subject Identification	98.62	90.96	68.20
	TEMP	Stress	99.00	94.88	90.00
		Subject identification	97.88	91.22	51.23
	ALL	Stress	99.86	95.56	90.00
		Subject identification	98.56	92.00	47.00
VERBIO	ACC	Stress	90.61	85.48	82.18
		Subject Identification	86.42	84.40	63.60
	BVP	Stress	92.96	88.95	85
		Subject Identification	85.63	80.86	48.64
	EDA	Stress	93.16	92.46	89.24
		Subject identification	88.30	83.83	45.12
	TEMP	Stress	92.64	90.42	86.32
		Subject identification	85.34	82.56	39.59
	ALL	Stress	95.00	93.96	90.94
		Subject	90.00	88.92	38.46

		Identification			

D. Comparison with SOTA

The proposed approach is also compared with existing approaches employed for stress recognition in FL settings, and the results are reported in Table 10. Most similar works have mainly focused on the benefits of FL in affective computing by generating data clients from unique datasets (i.e., as employed in the FL framework with standard benchmarks, namely MNIST, CIFAR-10, and CIFAR-100).
To the best of our knowledge, the number of studies using physiological signals in FL settings is relatively low compared to other case studies in affective computing, such as speech and facial expression-based studies. According to Table 10, we can see that existing works employed only FedAvg with the WESAD dataset and achieved good performance, as confirmed by our previous experimental section.
Consistent with compared works, for both databases, WESAD and VERBIO, the obtained performance behaves similarly.

Unlike the compared works [17,23,24,26], which analyze privacy concerns in stress recognition, we introduced a multitask model enabling us to strike a balance between the utility and privacy. This is achieved by selectively adding noise to layers prone to subject information leakage. The comparison provided in Table 4 reveals that prior studies achieved a performance comparable to our case study when employing FedAvg on WESAD or a private dataset. Through our experiments, we provided a comprehensive investigation of how to balance privacy preservation and stress recognition. The obtained results demonstrate that the combination of Multitask FL and DP can maintain good performance without sacrificing the user's privacy.

Table 4. Comparison of affect recognition studies using federated learning and privacy-preserving approaches. The following table reports the main approaches have been applied for stress detection by using physiological dataset.

Study	Dataset	FL algorithm	Data split	Accuracy
Almadhor et al. [23]	WESAD	FedAvg+Logistic Regresson	NA	85.75
Fauzi et al. [24]	WESAD	FedAvg+DNN network	NA	99.1
Can and Ersoy [17]	Private dataset	FedAvg+MLP	NA	88.55
Lee et al. [26]	WESAD	FedAvg+MLP	LOOCV	75.00
Our previous study [28]	WESAD	FedAvg+1DCNN+DP	5-fold CV	90.00
Our current study	VERBIO	FedAvg+1DCNN+DP	5-fold CV	88.67

Conclusion

The overall objective of our work was to present an approach to stress recognition that

ensures robustness while protecting the user's privacy. To this end, we developed a personalized multi-task federated model framework with differential privacy. We employed a user-level DP mechanism by injecting an amount of noise into personalized layers for perturbing identity while preserving task-specific utility. Usind multi task and DP (epsilon=15), we obtained 90% and 87% accuracy for recognizing emotions while limiting the re-identification accuracies to 47% and 38% on WESAD and VERBIO respectively. We extensively tested different parameters including layer of neural network, privacy budget, different noise distributions. As expected, adding noise to upper layer decrease the affect recognition performance less when compared to the last task specific layers. We also demonstrated the effect of data distribution to the performance. We believe that our results will guide researchers in determining suitable parameters and distributions for achieving the desired tradeoff between utility and privacy for affective computing applications. Currently, new gradient-based unsupervised adversarial attackers are attacking deep neural classification models to infer the privacy of distributed training gradient. In future works, to address this threat, we are planning to conduct additional experiments with federated differentially private generative adversarial networks that can provide better privacy protection and data diversity for widespread applications of physiological computing systems.

Authors' Contributions

M. Benouis and Y. S. Can together conceptualized and designed the paper and interpreted the results. M. Benouis performed coding, data processing and analysis. E. André supervised the study, provided valuable suggestions for improvement and critically revised the manuscript.

Acknowledgements

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

Conflicts of Interest

None declared.

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. 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.
8. 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.
9. 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.
10. 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.
11. 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.
12. V. Smith, C.-K. Chiang, M. Sanjabi, and A. S. Talwalkar, “Federated MTL,” Advances in neural information processing systems, vol. 30, 2017.
13. 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.
14. 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.
15. 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.
16. 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.
17. 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.
18. 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.
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. Almadhor, Ahmad, et al. "Wrist-based electrodermal activity monitoring for stress detection using federated learning." Sensors 23.8 (2023): 3984.
23. Fauzi, Muhammad Ali, Bian Yang, and Bernd Blobel. "Comparative analysis between individual, centralized, and federated learning for smartwatch-based stress detection." *Journal of Personalized Medicine* 12.10 (2022): 1584.

24. Lee, Yongho, et al. "Privacy preserving stress detection system using physiological data from wearable device." *Intelligent Human Systems Integration (IHSI 2023): Integrating People and Intelligent Systems* 69.69 (2023)
25. Wang, Meng, et al. "Identifying personal physiological data risks to the Internet of Everything: the case of facial data breach risks." *Humanities and Social Sciences Communications* 10.1 (2023): 1-15.
26. Can, Yekta Said, Bhargavi Mahesh, and Elisabeth André. "Approaches, applications, and challenges in physiological emotion recognition—a tutorial overview." *Proceedings of the IEEE* (2023).
27. A. El_Rahman, Sahar, and Ala Saleh Alluhaidan. "Enhanced multimodal biometric recognition systems based on deep learning and traditional methods in smart environments." *Plos one* 19.2 (2024): e0291084.
28. M. Benouis, Y. S. Can and E. André, "A Privacy-Preserving Multi-Task Learning Framework for Emotion and Identity Recognition from Multimodal Physiological Signals," 2023 11th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW), Cambridge, MA, USA, 2023, pp. 1-8.